

高品質音声分析合成を目的とした 音響パラメタ推定の研究

平成20年2月

和歌山大学大学院システム工学研究科

森勢 将雅

目次

第 1 章 序論	1
1.1. まえがき	1
1.2. デジタル信号処理によるスペクトル解析	2
1.2.1 音声のスペクトル推定	3
1.3. デジタル信号処理と音声	4
1.4. 音声分析合成法に関する先行研究	6
1.4.1 本論文で解くべき課題	6
1.4.2 スペクトル包絡の推定	9
1.4.3 Vocoder の問題点	10
1.5. 高品質音声分析変換合成法 STRAIGHT	11
1.5.1 STRAIGHT の問題	12
1.6. 音声の伝達関数に関する研究	12
1.6.1 音声放射に関する課題	13
1.7. 本研究の構成	13
第 2 章 分析時刻に依存しない周期信号のパワースペクトル推定法	17
2.1. はじめに	17
2.2. 周期信号を時間周波数分析した場合に起こる問題	18
2.2.1 周期信号のスペクトル	18
2.2.2 窓関数による切り出しがスペクトルに与える影響	20
2.3. 分析時刻に依存しない周期信号のパワースペクトル推定法: TANDEM	25
2.3.1 1つの窓関数による分析時刻依存性	25
2.3.2 2つ目の窓関数による干渉の抑圧	27
2.3.3 TANDEM による時間周波数分析	28
2.4. 評価	29
2.4.1 TANDEM に適した窓関数の候補	30
2.4.2 TANDEM 窓による性能の違い	31
2.4.3 単一の窓を用いた短時間フーリエ変換との比較	33
2.4.4 STRAIGHT における相補的時間窓との比較	35
2.4.5 TANDEM の耐雑音性の評価	36
2.4.6 基本周波数の誤差が時間変動に与える影響	37
2.4.7 音声分析に適した窓関数についての考察	41
2.5. まとめ	43

第 3 章 音声を用いたクロススペクトル法によるインパルス応答推定法	47
3.1. はじめに	47
3.2. インパルス応答の推定手法	48
3.2.1 音声の収録	52
3.2.2 インパルス応答の測定手順	52
3.2.3 多数回測定することによる特性のばらつき	54
3.2.4 離散系で実装することによる誤差と、誤差の補償	58
3.3. 評価	61
3.4. まとめ	62
第 4 章 音声を用いたクロススペクトル法に適した時間窓の検討	65
4.1. はじめに	65
4.2. クロススペクトル法を用いたインパルス応答推定と時間窓による振幅変形	66
4.3. 計算機シミュレーションによる評価	67
4.3.1 推定対象とするインパルス応答	67
4.3.2 ホワイトノイズ、音声との対比	67
4.3.3 音声と同様の振幅周波数特性を持つ雑音、ピンクノイズによるシミュレーション	70
4.3.4 ピーク間隔とダイナミックレンジを任意に設定した雑音によるシミュレーション	72
4.3.5 ダイナミックレンジに起因する推定誤差	74
4.3.6 音声を測定用信号としたインパルス応答推定に適した時間窓	75
4.4. まとめ	76
第 5 章 暗騒音と高調波歪の影響を軽減できるインパルス応答測定用信号の検討	81
5.1. はじめに	81
5.2. インパルス応答測定法と問題点	83
5.2.1 直接法	83
5.2.2 相互相関法	84
5.2.3 代表的な測定用信号	84
5.2.4 相互相関法の特徴と問題点	86
5.3. Warped-TSP の定義と特徴	87
5.3.1 定義式	88
5.3.2 Warped-TSP の性質	89
5.3.3 高調波歪の影響	89
5.3.4 直線状畳み込みにおける環状シフト量	92
5.3.5 測定環境に適したパラメタの設定法	93
5.3.6 Warped-TSP のまとめ	95
5.4. 実環境における実験	96
5.4.1 収録条件	96
5.4.2 測定環境に適した接続周波数の設定	97

5.4.3	測定用信号の評価指標	97
5.4.4	評価結果	99
5.4.5	直線状畳み込みについての検討	101
5.4.6	高調波歪の除去に関する実験	101
5.5.	まとめ	103
第 6 章 結論		109
参考文献		115
付録		128
A.	基本周波数の推定に関する先行研究	128
B.	相補的時間窓	129
C.	本論文で用いた公式の導出	130
謝辞		133

目 次

1.1	スペクトル推定法に関する手法の位置付け. 本論文で提案する方法との関係.	5
1.2	マイクロフォンで観測した有声音に含まれる情報. 声帯振動の周期信号 $x(t)$ とその基本周期 (基本周波数の逆数), 口の形や舌の位置により特定の周波数が強調されることを表現する調音フィルタ $V(\omega)$, 口元からマイクロフォンまでの伝達関数 $H(\omega)$. なお, 摩擦音の場合, $x(t)$ は雑音となる.	7
1.3	Vocoder の構成. スペクトル包絡には, 調音フィルタ $V(\omega)$, 音源の特性 (有声音の場合 $G(\omega)$) や, 図 1.2 に示す伝達関数まで含まれる.	8
2.1	周期信号の周波数特性. 任意のインパルス応答 (左上) を時間軸で周期的に配置した信号 (右上) のスペクトルは, 元のインパルス応答のスペクトル (左下) を周波数軸上で離散化したもの (右下) となる.	19
2.2	窓関数長を周期の 2 倍に設定した Hanning 窓のスペクトル.	20
2.3	基本周期 10 ms のパルス列 (上) と窓関数で切り出した波形 (中), 切り出された波形のパワースペクトル (下).	22
2.4	分析時刻を 5 ms シフトした場合の波形 (上), 窓関数で切り出された波形 (中), 切り出された波形のパワースペクトル (下).	23
2.5	パルス列の時間周波数分析. 時間周波数方向に周期的な谷が認められる.	24
2.6	TANDEM 窓 (上段) と周期 T_0 の周期信号の例 (下段). TANDEM 窓とは, 任意の窓関数を周期の半分だけシフトした 2 つの窓関数のことである.	26
2.7	TANDEM を用いたパルス列の時間周波数分析例. TANDEM は時間方向で周期的に現れた谷が取り除き, 分析時刻によらず安定したパワースペクトルを与える. TANDEM におけるパワースペクトルの周波数変動は ω_0 を周期とする振動として現れる.	29
2.8	TANDEM により得られたパワースペクトルを周波数方向に平滑化して得られたスペクトログラム. 時間周波数方向に滑らかな応答が得られる.	30
2.9	3 種類の窓関数の時間波形.	31
2.10	3 種類の窓関数のパワースペクトル.	32
2.11	単一の窓関数を用いた短時間フーリエ変換を用いた場合との比較. Hanning 窓の場合.	34
2.12	単一の窓関数を用いた短時間フーリエ変換を用いた場合との比較. Blackman 窓の場合.	35
2.13	単一の窓関数を用いた短時間フーリエ変換を用いた場合との比較. Nuttall 窓の場合.	36

2.14	単一の各窓関数から求めた時間変動の比較. 単一の窓関数を用いた短時間フーリエ変換では, 窓の持続時間と時間変動とがほぼ一致する.	37
2.15	窓関数の持続時間と時間変動との関係. STRAIGHT と比較しても, 持続時間, 時間変動ともに優れた窓が存在する.	38
2.16	窓関数の種類・長さとは時間変動との関係. SNR60 dB における結果.	39
2.17	窓関数の種類・長さとは時間変動との関係. SNR50 dB における結果.	39
2.18	窓関数の種類・長さとは時間変動との関係. SNR40 dB における結果.	40
2.19	窓関数の種類・長さとは時間変動との関係. SNR30 dB における結果.	40
2.20	窓関数の種類・長さとは時間変動との関係. SNR20 dB における結果.	41
2.21	基本周波数の推定誤差が TANDEM 窓の持続時間に与える影響. Hanning 窓の場合.	42
2.22	基本周波数の推定誤差が TANDEM 窓の持続時間に与える影響. Blackman 窓の場合.	43
2.23	基本周波数の推定誤差が TANDEM 窓の持続時間に与える影響. Nuttall 窓の場合.	44
2.24	基本周波数に 10 % の誤差含まれる場合における TANDEM 窓の持続時間と時間変動.	45
3.1	クロススペクトル法における窓掛け.	50
3.2	音声を入力とした場合のクロススペクトル法.	51
3.3	マイクの設置位置について.	53
3.4	収録音声と暗騒音の 1/3 オクターブバンド毎のエネルギー.	54
3.5	クロススペクトル法により測定されたインパルス応答.	55
3.6	低域の補正処理を行ったインパルス応答.	55
3.7	測定されたインパルス応答の振幅周波数特性 (a) と, 低域の補正処理, 反射波の抑制を行ったインパルス応答の振幅周波数特性 (b).	56
3.8	エネルギーを正規化したインパルス応答 (処理後) の振幅周波数特性 (a) と, パワースペクトルで重み付けした群遅延特性 (b) の, 平均 (実践) と標準偏差 (破線).	57
3.9	線形の時間軸で観測したインパルス応答.	59
3.10	非線形伸縮された時間軸で観測したインパルス応答.	59
3.11	時間軸伸縮処理前後の応答から求めた振幅周波数特性の誤差.	60
3.12	補償前 (a) と補償後 (b) の誤差信号.	61
3.13	振幅周波数特性の平均 (実線) と標準偏差 (破線). 上段は処理前, 下段が処理後.	63
3.14	パワースペクトルで重み付けした群遅延特性の平均 (実践) と標準偏差 (破線). 上段は処理前, 下段が処理後.	64
4.1	推定対象となるインパルス応答の時間波形.	68
4.2	推定対象となるインパルス応答の振幅周波数特性.	68
4.3	インパルス応答推定に用いた音声の振幅周波数特性.	69

4.4	正規化誤差エネルギーと時間窓長 N とインパルス応答長 M の比との関係。 測定用信号はホワイトノイズ.	70
4.5	正規化誤差エネルギーと時間窓長 N とインパルス応答長 M の比との関係。 測定用信号は音声.	71
4.6	正規化誤差エネルギーと時間窓長 N とインパルス応答長 M の比との関係。 測定用信号は音声と同じ振幅周波数特性を持つ雑音.	72
4.7	正規化誤差エネルギーと時間窓長 N とインパルス応答長 M の比との関係。 測定用信号はピンクノイズ.	73
4.8	全極型の伝達関数を用いて整形した測定用信号の振幅周波数特性.	74
4.9	正規化誤差エネルギーと時間窓長 N とインパルス応答長 M の比との関係。 測定用信号はダイナミックレンジ 10 dB の雑音.	75
4.10	正規化誤差エネルギーと時間窓長 N とインパルス応答長 M の比との関係。 測定用信号はダイナミックレンジ 20 dB の雑音.	76
4.11	正規化誤差エネルギーと時間窓長 N とインパルス応答長 M の比との関係。 測定用信号はダイナミックレンジ 30 dB の雑音.	77
4.12	正規化誤差エネルギーと時間窓長 N とインパルス応答長 M の比との関係。 測定用信号はダイナミックレンジ 40 dB の雑音.	78
4.13	正規化誤差エネルギーと時間窓長 N とインパルス応答長 M の比との関係。 測定用信号はダイナミックレンジ 50 dB の雑音.	78
4.14	正規化誤差エネルギーと時間窓長 N とインパルス応答長 M の比との関係。 測定用信号はダイナミックレンジ 60 dB の雑音.	79
4.15	ダイナミックレンジと正規化誤差エネルギーとの関係. $N/M = 16$ について.	79
5.1	直接法によるインパルス応答測定の流れ図.	84
5.2	相互相関法によるインパルス応答測定の流れ図.	85
5.3	Warped-TSP(上図) と逆 Warped-TSP(下図) の時間波形(左図) と振幅周波数特性(右図).	90
5.4	Warped-TSP における, 接続周波数による振幅周波数特性の変化. 接続周波数は, 125 Hz から 8000 Hz までオクターブ毎に設定している. また, 各信号が等しいエネルギーを持つようにしている.	91
5.5	Warped-TSP の第 2 次・第 3 次高調波歪 ($\tau_{2,h}$ and $\tau_{2,l}$, $\tau_{3,h}$ and $\tau_{3,l}$). 高調波歪は太線の垂線で示される主応答とは時間的に重ならない.	93
5.6	定義式に基づいて作成した Warped-TSP(上図) と, 適切な環状シフトを行った Warped-TSP(下図).	94
5.7	様々な場所で測定した暗騒音の 1/3 オクターブバンド毎のエネルギー.	95
5.8	測定環境 1 における収録条件(簡易防音室・ヘッドホン測定)	97
5.9	測定環境 2 における収録条件(防音室・スピーカー測定)	98
5.10	各測定環境における接続周波数と雑音低減比の関係および適切な接続周波数の設定	99
5.11	暗騒音区間の定義.	100

5.12	測定環境1(図5.8)の暗騒音(a)と、各測定用信号(b)の、1/3オクターブバンド毎のエネルギー．	101
5.13	測定環境2(図5.9)の暗騒音(a)と、各測定用信号(b)の、1/3オクターブバンド毎のエネルギー．	102
5.14	測定に用いたLin-TSPのスペクトログラム	103
5.15	測定に用いたLog-TSPのスペクトログラム	104
5.16	測定に用いたWarped-TSPのスペクトログラム(測定環境1)	105
5.17	測定に用いたWarped-TSPのスペクトログラム(測定環境2)	105
5.18	測定環境1(a)，および測定環境2(b)における，同期加算した場合(白)としない場合(黒)の各測定用信号によるSNR．	106
5.19	環状畳み込み(黒)と直線状畳み込み(白)を用いた場合のSNR．	106
5.20	Lin-TSPを用いた測定で得られたインパルス応答と高調波歪のスペクトログラム．	107
5.21	Log-TSPを用いた測定で得られたインパルス応答と高調波歪のスペクトログラム．	107
5.22	Warped-TSPを用いた測定で得られたインパルス応答と高調波歪のスペクトログラム．	108

表 目 次

2.1	実験条件	32
5.1	各 TSP の特徴	96
5.2	測定条件 (測定環境 1)	98
5.3	測定条件 (測定環境 2)	99

第 1 章 序論

1.1. まえがき

入力信号から目的とする情報を抽出する、あるいは不要な情報を取り除く等の信号処理技術は、古くから研究されてきた。信号処理は、本論文で対象とする音声以外にも、心電図、音響信号、電波、超音波やデータ通信、符号化、暗号化などに応用される重要な技術である。初期の信号処理は、アナログ信号を対象に行われており、アナログ信号をアナログ回路により処理するものであった¹⁾。

これに対し、近年では、デジタル信号処理が普及している。初期のデジタル信号処理は、アナログ信号処理の技術を近似するものとされてきた¹⁾。しかし、実際には、デジタル信号処理は、アナログ信号処理を近似するだけではなく、アナログ信号処理では困難な暗号化や符号化といった技術を可能としている。AD 変換により得られるデジタル信号は、アナログ信号と比べ、環境変化やメディアの劣化による影響を受けにくく、安定した情報を保存できるというメリットもある⁴⁾。さらに、デジタル信号はアナログ信号と標本化定理で保障される帯域において相互変換が可能である¹。すなわち、デジタル信号処理は、アナログ信号処理を近似するための分野ではなく、アナログ信号処理では困難な信号処理をも可能とする分野と位置付けられる。

しかし、デジタル信号処理では、処理対象のアナログ信号を AD 変換する必要がある。また、スペクトル解析を行う離散フーリエ変換も計算量が多い。これらの問題点から、デジタル信号処理技術は有用ではあるものの、実用的ではないとされていた。1965 年に高速フーリエ変換 (Fast Fourier transform: FFT)^{5, 6)} が提案され、デジタル信号処理における計算量の問題点が解決された。これにより、デジタル信号処理は急速に普及し、数多くの信号処理技術が提案されることとなる。

¹ ただし、標本化定理は、現実には実現不可能な理想フィルタの存在を仮定している。また、振幅の量子化による誤差が含まれるため、現実的には完全に復元することはできない。標本化定理に関する研究も行われており^{2, 3)}、入力とするアナログ信号が、AD/DA 変換後に受ける影響についても論じられている。

1.2. デジタル信号処理によるスペクトル解析

確率過程のパワースペクトル推定法^{10, 11, 8)}は、様々なデジタル信号処理の基礎となる重要な手法である。パワースペクトル推定法は、主に AR モデル⁸⁾のようにモデルに基づく方法と、フーリエ変換に基づいて解析する方法とが提案されている。さらに、フーリエ変換を用いる方法では、自己相関を窓関数で処理した信号に対しフーリエ変換を行う Blackman-Tukey 法と、信号そのものにフーリエ変換を行うピリオドグラム法とに分けられる¹²⁾。

最も簡単なデジタル信号のパワースペクトル推定は、推定対象となる信号に対しフーリエ変換を行い、スペクトルの振幅を2乗することである。この分析信号からフーリエ変換によりパワースペクトルを推定する方法は、ピリオドグラム法と呼ばれる。Blackman-Tukey 法では、信号を直接フーリエ変換せず、自己相関をフーリエ変換することで直接パワースペクトルを求めている。Blackman-Tukey 法は、信号の自己相関のフーリエ変換がパワースペクトルと等価であるという Wiener-Khintchine の定理に基づく方法と位置付けられる。

Welch 法⁷⁾は、定常信号を窓関数の分析時刻をシフトさせながら切り出し、全てのフレームについて得られたパワースペクトルを平均し、最終的なパワースペクトルを得る方法である。Welch 法で用いる、信号を特定の区間について切り出して行うフーリエ変換は、短時間フーリエ変換¹³⁾という。短時間フーリエ変換は、パワースペクトル推定法や時間周波数解析¹⁴⁾の一部でもあるスペクトログラムにも用いられる基礎的な分析方法である。Welch 法は、各フレームのパワースペクトルの分散を2次の確率過程について求めている。また、全てのフレームのパワースペクトルを平均することで分散が低減でき、安定したパワースペクトルが得られることを示している。Welch 法は、波形のパワースペクトルを平均することで分散を低減できることから、ピリオドグラム法により得られるパワースペクトルの分散を低減できる推定法と位置づけることができる。

入出力間に存在する伝達関数を推定する方法としてクロススペクトル法⁹⁾が提案されている。クロススペクトル法では、フレームを平均することで分散が低減できる Welch 法の性質を利用し、伝達関数の推定精度を向上している。また、窓関数のシフト量を窓関数長の半分程度 (Hanning 窓でおよそ 63 %) オーバーラップさせることで平均回数を増やし、推定精度を向上できることが示されている¹⁵⁾。Welch 法の性質を利用することで、伝達関数に含まれる雑音成分のパワー比率を推定することもできる。このパワー比率は、コヒーレンス関数⁹⁾といわれ、システムに含まれる雑音成分の指標として用いられる。コ

ヒーレンス関数とクロススペクトル法を利用することで、伝達関数の特性と伝達関数に含まれる雑音成分が推定できる。このように、Welch 法は、確率過程のパワースペクトル推定だけではなく、伝達関数やフィルタの特性、システムに含まれる雑音のパワーを推定する方法にも応用される。

短時間フーリエ変換は、長時間の信号を短い区間について切り出して個別にフーリエ変換を行う方法と位置付けられる。長時間の信号から短い区間を切り出す操作は、波形と切り出す区間の長さを持つ矩形窓との積として表現される。窓関数による処理の影響は、スペクトル領域で窓関数のスペクトルと系列のスペクトルを畳み込み演算したものとして現れる。推定対象となる信号のパワースペクトルに大きなパワーと小さなパワーとが隣接した周波数に存在する場合、メインローブの広い窓関数では小さなパワーが観測できないという問題が生ずる。この問題に対処することを目的として窓関数の設計に関する研究が進められた。メインローブが狭く、サイドローブの小さい窓関数として Hanning 窓や Blackman 窓など多数の窓関数が提案されている^{16, 17)}。窓関数の設計法に関しても、パワースペクトル推定精度を向上させる方法と位置付けられる。ただし、窓関数がパワースペクトル推定に与える影響については、確率過程を前提にせずとも成立する。

近年の研究は、これらの手法に基づき、長時間では無相関とはならない複雑な信号の分析へと応用されていくこととなる。次節では、本論文で対象とする音声のスペクトル推定について述べる。

1.2.1 音声のスペクトル推定

声帯振動を有する音声は、駆動音源に周期が存在するため、大きな相関を有する信号となる。周期を有する信号の解析では、周期性の存在がパワースペクトルの推定結果に大きな影響を与える。そのため、音声のスペクトル推定を行う場合、確率過程で生じた分散の影響だけではなく、周期性の存在がパワースペクトルに与える影響を低減させる必要がある。周期性を有する信号のスペクトル推定法としてピッチ同期分析⁵⁷⁾が提案されている。ピッチ同期分析は、切り出しに用いる窓関数の長さをピッチ(基本周波数)をパラメタとして決定することにより、有声音のスペクトル推定精度を向上させる方法である。しかし、ピッチ同期分析を用いた場合においても、得られるパワースペクトルは、スペクトルの干渉により、窓関数の分析位置に依存して大きく異なる問題があった。

この周期性が原因で生ずるスペクトルの干渉については、窓関数により信号を切り出す限り生じる本質的な問題とされていた。この問題には、線形予測^{32, 33, 34)}を用いる推

定法や、ケプストラム^{25, 26)}に基づく推定法によって取り除く試みが行われてきた。線形予測は、ARモデルに基づく、モデルを用いたパワースペクトル推定法に位置付けられる。ケプストラムは、音声のスペクトルが、周期性により生ずるスペクトル成分と求めるべきスペクトル成分との積になることに着目している。音声信号のケプストラムでは、高次のケフレンシーにスペクトル干渉の成分が集中する。ケプストラムから低次の成分のみを切り出すリフタリング処理により、干渉の影響を低減させている。しかし、これらの方法については1.4.2節で詳しく述べるが、スペクトルの干渉を直接的に低減させる方法とはならない。

1997年にスペクトルの干渉を取り除く方法として、STRAIGHT^{40, 41, 42)}が提案された。STRAIGHTでは、1つの窓関数により生じる干渉成分を、2つ目の特殊な窓関数のパワースペクトルを用いて相殺することでスペクトルの干渉成分を低減している。すなわち、STRAIGHTは、ケプストラムや線形予測とは異なりスペクトルの干渉を直接低減するための方法に位置づけられる。また、STRAIGHTでは、ピッチ同期分析同様に窓関数長を基本周波数をパラメタとして決定する。このことから、STRAIGHTはピッチ同期分析を改良した方法ともいえる。STRAIGHTについては1.5節で詳しく説明する。

本論文では、周期信号のスペクトル推定精度を向上させることを目的とし、スペクトルの干渉を低減する新たな方法TANDEMの提案を行う。TANDEMは、Welch法のように複数の窓関数により切り出した波形のパワースペクトルを平均する方法である。ただし、Welch法とは異なり、窓関数により切り出す時刻を基本周波数に依存させることにより、スペクトルの干渉をSTRAIGHT以上に低減するための方法である。すなわち、TANDEMは、STRAIGHTを改良したスペクトル推定法という位置付けとなる。

これまでに示したスペクトル推定法と本論文で提案する方法は図1.1のような位置付けになる。

1.3. デジタル信号処理と音声

最も簡単な音声モデルは、音源を発生するモデルと、声道形状や舌の位置により任意の音韻性を与えるフィルタから構成される¹⁸⁾。音声には、様々な種類があり、種類に応じて駆動音源が異なる。例えば、有声音における音源は周期的な声帯振動であり、摩擦音における音源は雑音である。このような音源を、任意の音韻性を与えるフィルタに通すことで音声信号が出力される。マイクロフォンで観測された音声信号には、これらの情報に加え、口元から観測点までにおける伝達関数が含まれる。伝達関数は、室内におけ

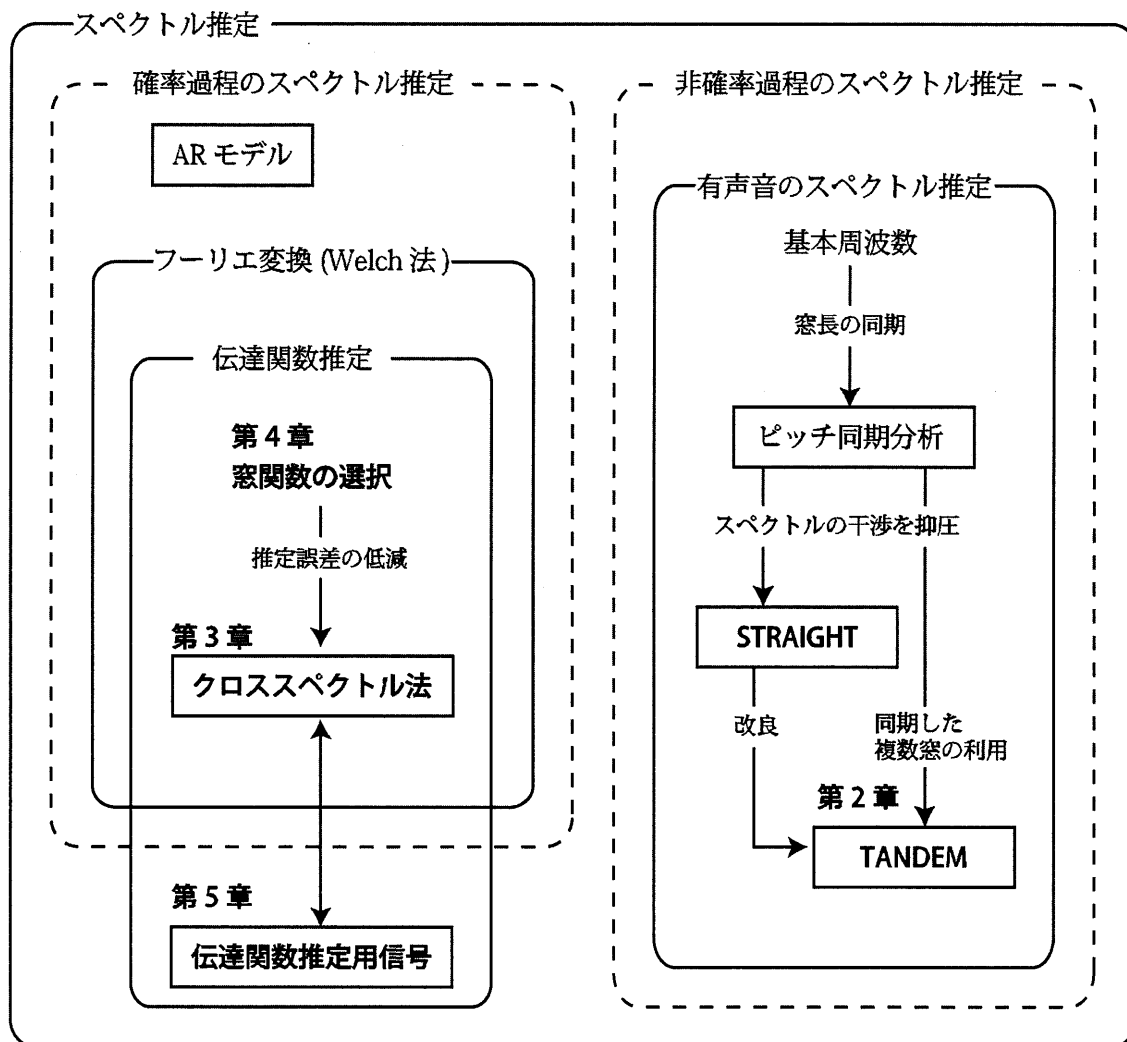


図 1.1 スペクトル推定法に関する手法の位置付け. 本論文で提案する方法との関係.

る反射や残響の影響を記述したスペクトル情報である. 例えば有声音に関すると, 図 1.2 のような情報が含まれることとなる.

図 1.2 に示すように, マイクロフォンにより収録した有声音 $y(t)$ には, 声帯の振動周波数 (以下では基本周波数とする), 声帯振動の波形情報, 口の形や舌の位置により特定の周波数が共鳴するフィルタ $V(\omega)$, 口元から観測位置までの伝達関数 $H(\omega)$ が含まれる. 音声を対象としたデジタル信号処理では, これらの音響パラメタを高精度に推定することが第一の課題となる. 高品質音声合成に必要な音響パラメタは, 声帯振動波形, 調音フィルタ $V(\omega)$, 基本周波数である. 伝達関数は, 室内の反射, 残響といった音声の明瞭

度を損なわせるスペクトル情報であり、高品質音声合成には不要な除去すべき音響パラメタである。

1.4. 音声分析合成法に関する先行研究

本研究は、高品質音声分析合成法の提案を目指して進めてきた。音声合成を目的とした音声分析では、音声から基本周波数とスペクトル情報(以下ではスペクトル包絡とする)を取り出す信号処理を行う。音声合成では、基本周波数とスペクトル包絡から音声を合成する。このように音声を音響パラメタに分解し、得られたパラメタを用いて再合成する方式は Vocoder^{19, 20)} と呼ばれている。Vocoder の概略図を図 1.3 に示す。図 1.3 より、スペクトル包絡は、音声の種類により異なる情報を含むことがわかる。無声音におけるスペクトル包絡は、調音フィルタ $V(\omega)$ と口元から観測点までの伝達特性 $H(\omega)$ である。有声音におけるスペクトル包絡では、これらのスペクトル情報に加え、声帯振動波形のスペクトル情報 $G(\omega)$ も含まれる。スペクトル包絡に関しては、図 1.3 に示す調音フィルタ $V(\omega)$ だけではなく、声帯振動波形のスペクトル情報 $G(\omega)$ や伝達関数 $H(\omega)$ までを含む、1つのスペクトルの推定が行われてきた。

1.4.1 本論文で解くべき課題

高品質音声合成には、基本周波数とスペクトル包絡を推定することが必要である。基本周波数は、声帯振動回数という数値であり、人間が知覚する声の高さにも大まかに対応している。そのため、音声を高品質に合成するためには、基本周波数を正確に推定することが要求される。リアルタイムで声を変換するようなシステムでは、高精度かつ高速に基本周波数を推定する必要がある。実環境では収録された音声波形に雑音が含まれるため、雑音に対しても頑健に動作する方法が望まれる。従来法は、これらの要望を満たすために研究されてきた。具体的な基本周波数推定法の先行研究については付録にて示す。

図 1.3 に基づいて有声音が構成されているとすると、有声音は、スペクトル包絡を逆フーリエ変換することで得られる時間波形を、時間軸上に基本周期の間隔で配置したものとなる。Vocoder の構成図(図 1.3)を考えると、声帯振動のスペクトル情報はスペクトル包絡に含まれる。そのため、本来声帯振動の連なりである駆動音源を、任意の時刻 τ で生ずるパルス $\delta(t - \tau)$ の連なりとできる。基本周期を T_0 とした場合、駆動音源は周期 T_0

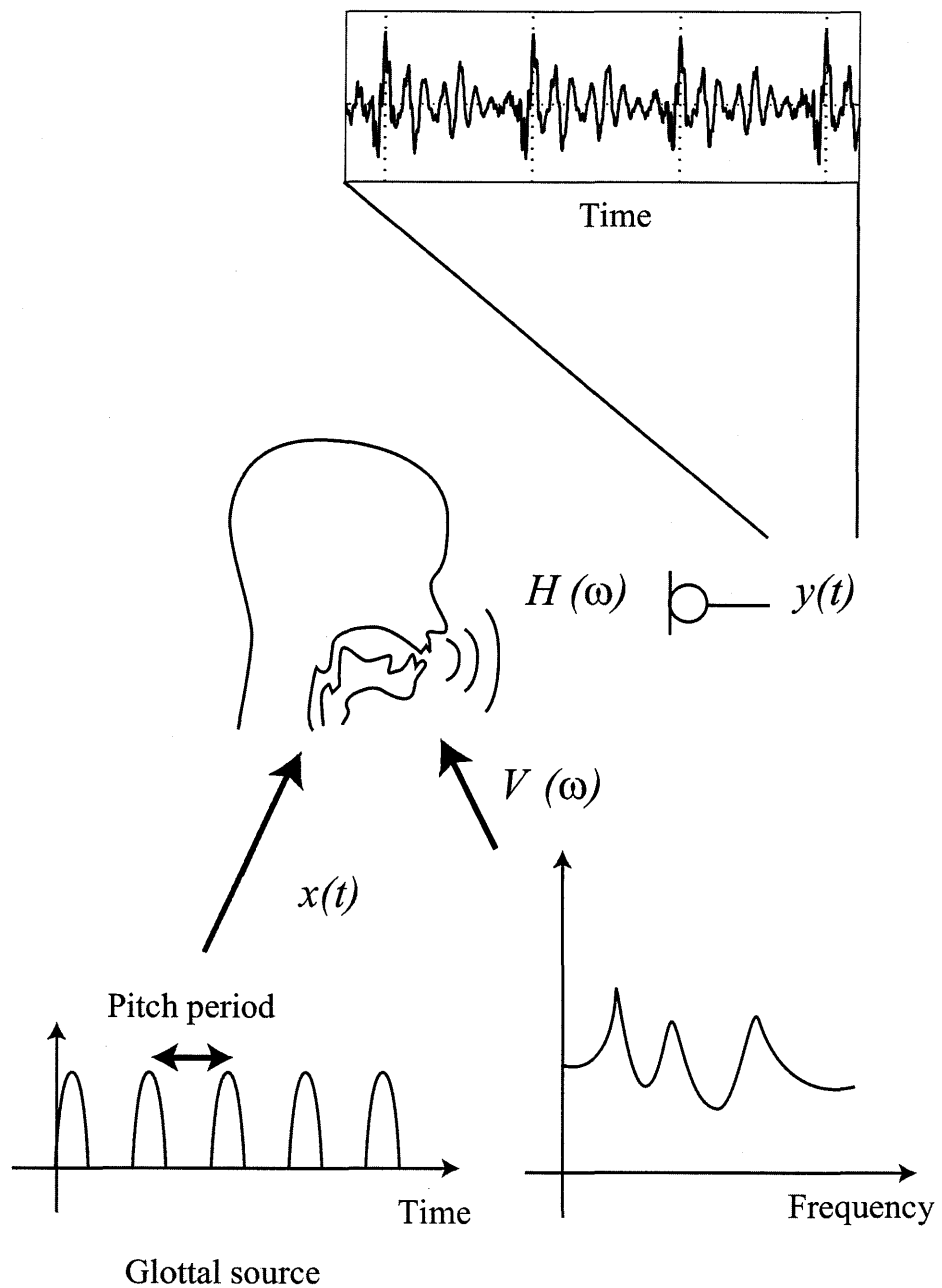


図 1.2 マイクロフォンで観測した有声音に含まれる情報. 声帯振動の周期信号 $x(t)$ とその基本周期 (基本周波数の逆数), 口の形や舌の位置により特定の周波数が強調されることを表現する調音フィルタ $V(\omega)$, 口元からマイクロフォンまでの伝達関数 $H(\omega)$. なお, 摩擦音の場合, $x(t)$ は雑音となる.

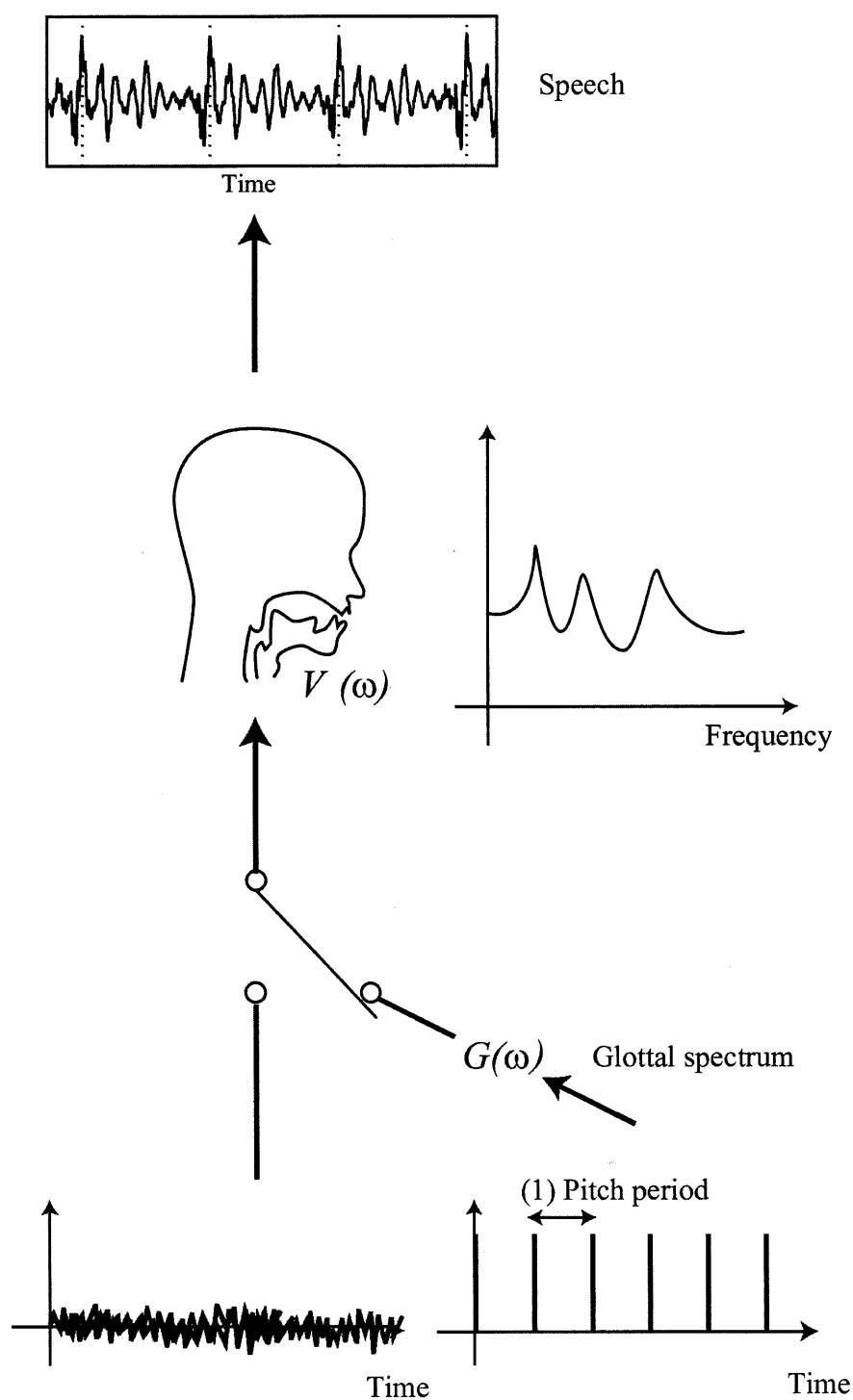


図 1.3 Vocoder の構成. スペクトル包絡には, 調音フィルタ $V(\omega)$, 音源の特性 (有声音の場合 $G(\omega)$) や, 図 1.2 に示す伝達関数まで含まれる.

のパルス列となる．パルス列 $x(t)$ を下式で表現した場合，そのスペクトル $X(\omega)$ は以下のようになる．

$$x(t) = \sum_{n=-\infty}^{\infty} \delta(t - nT_0) \quad (1.1)$$

$$X(\omega) = \sum_{n=-\infty}^{\infty} \delta(\omega - n\omega_0)$$

$$\omega_0 = \frac{2\pi}{T_0}$$

式1.1 の関係の導出については付録で行う．スペクトル包絡 $P(\omega)$ により処理された音声信号 $y(t)$ とそのスペクトル $Y(\omega)$ は，以下のようになる．

$$y(t) = x(t) * p(t) \quad (1.2)$$

$$Y(\omega) = P(\omega) \sum_{n=-\infty}^{\infty} \delta(\omega - n\omega_0) \quad (1.3)$$

$$\omega_0 = \frac{2\pi}{T_0} \quad (1.4)$$

* は，畳み込みを表す記号である．このときスペクトル包絡 $P(\omega)$ は，有声音ならば $G(\omega)$, $V(\omega)$, $H(\omega)$ の積，それ以外ならば $V(\omega)$, $H(\omega)$ の積となる．なお，マイクロフォンを話者の口元に設置するような場合，伝達関数 $H(\omega)$ は実質的に無視できる．音声波形 $y(t)$ のスペクトル $Y(\omega)$ は，特定の周波数でのみ任意の振幅・位相を持つ．式1.3 より，値を有する周波数は，基本周波数の整数倍のみとなる．このように，任意の周期をもつパルス列を畳みこむ処理は，時間波形・スペクトルの両方に，影響を与えることとなる．

高品質音声分析合成の実現には，スペクトル包絡 $P(\omega)$ の推定が必要となる．また，音声の基本周波数変換やモーフィング⁴⁸⁾ といった加工に対する劣化を抑えるためには，音声の定常部分において，分析時刻に依存せず安定したパワースペクトルを与える必要がある．

1.4.2 スペクトル包絡の推定

スペクトル包絡は，Vocoder の実現に必要な主要な音響パラメタである．式1.3 より，音声波形におけるスペクトル包絡 $P(\omega)$ は離散化されており，基本周波数の整数倍でのみ値を持つことが分かる．そのため，スペクトル包絡を推定する問題は，離散化したスペクトルから元のスペクトルを近似的に求める問題となる．

ケプストラムに基づくスペクトル包絡推定

ケプストラム^{25, 26)}を用いることで、スペクトル包絡の推定が行える。ケプストラムは、音声の振幅スペクトルの対数を逆フーリエ変換することで得られるパラメタである。スペクトルを逆フーリエ変換したパラメタのため、ケプストラムは時間の次元をもつ。このときの時間軸はケフレンシーと呼ばれる。

音声波形は、式1.2のように、パルス列とスペクトル包絡との畳み込みとして表現できる。時間軸における畳み込みは積に、積の対数は和になることから、ケプストラムは、時間軸における畳み込みを和として表現していることになる。このように、畳み込みで表わされる2つの系列を和で表わせることにケプストラムの本質が存在する。

音声信号のケプストラムでは、高次のケフレンシーに周期的なパルス列に関する情報が集中することが分かっている。そのため、ケプストラムから低次の成分のみを切り出す(この処理をリフタリングと呼ぶ)ことで、スペクトル包絡を近似的に取り出すことが出来る。

線形予測に基づくスペクトル包絡推定

線形予測^{32, 33, 34)}は、音声波形の時系列と、線形予測値との間の2乗平均誤差が最小となるように線形予測係数を定める方法である。この線形予測係数から得られるスペクトルは、全極型となる。すなわち、入力となる音声に対し、全極型のフィルタを用いて、2乗平均誤差が最小となるパラメタ推定を行う方法といえる。スペクトル包絡が、声帯から口元を断面積の異なる音響管により構成されたとすると、スペクトルは全極型となる。ただし、実際の音声は鼻音など口以外からの放射も含まれるため、調音フィルタは一般的には極零型になる。

1.4.3 Vocoderの問題点

従来のVocoderとして、ケプストラムVocoder(ホモモルフィックVocoderともいう)^{36, 37, 38)}や、線形予測に基づくPARCOR Vocoder³⁹⁾が提案されている。この他にも、音声を帯域フィルタ群で分析し、各フィルタの振幅、位相を特徴量とするPhase Vocoderも提案されている。ただし、本論文では、基本周波数とスペクトル包絡を特徴量として合成する方式をVocoderと呼ぶこととする。

従来のVocoderでは、以下の理由により音質が劣化する。

- スペクトル包絡の近似精度
- 声帯振動をパルスで近似する，あるいはスペクトル包絡の時間波形をゼロ位相にすることで生ずる，機械的な音色
- 有声無声を切り替える際の不連続感
- 有声・無声判定，あるいは基本周波数推定の誤り

本論文では，スペクトル包絡の近似精度の向上を目的としている．

ケプストラムに基づく方法では，高次ケプストラムをリフタリングで処理することにより，スペクトル包絡を推定している．このとき，リフタリングにより切り出すケフレンシーの次数により，スペクトル包絡の情報量が異なる．極めて低いケフレンシーのみを取り出すようなリフタリング処理では，スペクトルが過剰に平滑化され，合成音声の品質が劣化する．一方高いケフレンシーまで取り出すようにリフタリングすると，パルス列のスペクトルとの分離が不十分となり，品質が劣化する．そのため，ケプストラムに基づいて得られるスペクトル包絡には，リフタリングの次数に依存した誤差が含まれることとなる．線形予測においても同様に，過剰な数の極は，パルス列のスペクトルによる影響が含まれるようになる．少なすぎる数の極では，十分な推定精度を得ることができない．

さらに，求めるべきスペクトル包絡の持つ情報量は，個人性・音韻により異なることが考えられる．スペクトル包絡を十分な精度で近似でき，かつパルス列のスペクトルの影響を分離できる条件が常に成立するとは限らず，推定されるスペクトル包絡には必ず誤差が含まれることとなる．

1.5. 高品質音声分析変換合成法 STRAIGHT

1997年に，Vocoderでありながら元音声に匹敵する品質の音声を合成することができる STRAIGHT^{40, 41, 42)} が提案された．STRAIGHTによる音声分析は，基本周波数推定法，スペクトル包絡推定法，非周期性指標^{43, 44, 45)} 推定法の3つから構成される．STRAIGHTで推定されるスペクトル包絡は，以下の2つの条件を満足している．

- 基本周波数の整数倍では，スペクトル包絡本来の値を持つ
- 分析時刻に依存しないスペクトル包絡を与え，周波数方向にも平滑化されている

STRAIGHTにより得られたスペクトル包絡に基づく合成音は極めて高品質であり⁴⁶⁾，声質の分析や変換などの研究が多数行われている^{47, 48, 49)}．

1.5.1 STRAIGHT の問題

しかし、現状の STRAIGHT には以下に示す問題がある。

- 時間周波数平滑化を行う処理の複雑さ、計算コスト、予め最適化した多数のパラメータが処理に必要となる実装の煩わしさ
- 計算時間が膨大な基本周波数推定法
- 破裂子音等の合成法が提案されていない
- データ量が音声波形の 50 倍程度

このように、より自然な音声合成や音声加工を行うためには、いくつかの課題を解決する必要がある。特に real-time STRAIGHT⁵⁰⁾ は、基本周波数推定法を別の方法に置き換えられて実装している。そのため、高品質のヴォイスチェンジャーを実現するためには、計算コストを削減する必要がある。STRAIGHT のデータ量に関しては、時間・周波数双方から知覚に影響を及ぼさない範囲で削減する検討が行われている^{51, 52)}。基本周波数推定は、本論文 2 章で提案する方法を用いることにより改善できる可能性があり、現在検討を進めている。破裂子音に関しては、STRAIGHT においても、他の子音と同様に雑音を音源として駆動している。しかし、破裂音は摩擦音とは異なる音源となるため、より高品質な音声を合成するためには、子音の種類に応じて駆動音源を変える必要がある。

本論文は、STRAIGHT スペクトルを推定する場合の問題である、処理の複雑さ、計算コストの解決を第一の目的とする。具体的には、STRAIGHT における 2 つの制約を、STRAIGHT より高精度に満たし、高速で簡便なアルゴリズムを提案する。

1.6. 音声の伝達関数に関する研究

図 1.2 に示すように、マイクロフォンで収録した音声には口元から観測点までの伝達関数が含まれる。伝達関数は、口元から観測点までの距離や、収録する室内環境により異なる。口元に設置したマイクロフォンで収録した音声の場合の伝達関数は、反射・残響の影響が直接音に比べ十分小さいものとなり、品質を妨げる原因とはならない。しかし、口元から離れた位置に設置したマイクロフォンでは、背景雑音や、室内の反射・残響の影響により品質が劣化する。伝達関数を求め、音声合成に不要な成分を取り除くことが可能ならば、合成音声の更なる品質向上が見込まれる。

伝達関数を音圧に関して推定する研究は、1960年代より始まっている。Flanagan は、1960年にマネキンの口内にスピーカを設置し、水平面上の音圧分布を測定した⁵³⁾。測定結果より、角度によって差があることが明らかとなっている。その後、マネキンではなく人間が発声した声や、ダミーヘッドを用いて放射の水平面上における音圧分布を測定する例が示されている^{54, 55)}。ダミーヘッドとは、人間の上半身を模擬したマネキンであり、口元にスピーカを、耳元にマイクロフォンを備えている。ダミーヘッドは口からの放射と頭部および体による反射・回折が考慮されている。

近年では、Welch 法をベースとしたクロススペクトル法を用いることで、人間の口から発せられる声の放射を位相情報まで含めて推定できる方法が提案されている⁵⁶⁾。しかし、特に高域において SNR が低下することが分かっており、10 kHz 以上の成分は低域通過フィルタにより除去する必要があるとされていた。一般的に人間が聴こえる周波数の上限とされる 20 kHz までの伝達関数を高精度に推定することができれば、音声合成の品質向上に貢献できる可能性がある。

1.6.1 音声放射に関する課題

口元から観測点までの伝達関数は、クロススペクトル法により推定することができる。クロススペクトル法では、入力信号として無相関な白色雑音をベースに議論されている。ダミーヘッドを用いることで音声を入力信号とした場合と、インパルス応答推定用信号でもある M 系列信号を入力信号とした場合とについて、比較が行われている⁵⁶⁾。比較の結果、音声のエネルギーが大きい帯域については、差が 1dB 以下となっていることが示されている。しかし、高域については、SNR の低下により安定した伝達関数が推定できないため、特に検討されていなかった。高域までの伝達関数の妥当性を示すためには、どちらの方法を用いて推定された伝達関数においても、全帯域について高い SNR であることが要求される。

本論文では、クロススペクトル法により得られた伝達関数より高域の SNR を向上させる後処理を提案する。同時に、高い SNR で伝達関数が推定できる測定用信号を開発する。

1.7. 本研究の構成

本論文の構成は以下の通りである。

第2章では、音声スペクトルから分析時刻に依存せず安定したスペクトル包絡を推定するためのアルゴリズムを提案する。まず、本章で簡単に説明した、音声スペクトルについて詳しい説明を行う。特に、音声分析を行う際必要となる窓関数による波形の切り出しについて説明し、窓関数による切り出しが音声スペクトルに与える影響を示す。周期信号である有声音を、有限の長さの窓で切り出したという条件の下で、本論文で解くべき問題を明らかにする。次いで、解決法となる周期信号分析法TANDEMを提案する。TANDEMが、STRAIGHTと同様に、分析時刻に依存せず、基本周波数の整数倍において、スペクトル包絡本来のパワースペクトルをもつことを示す。計算機シミュレーションによりTANDEMとSTRAIGHTとの比較を行い、TANDEMの有効性を示す。具体的には、分析時刻に依存して変動するエネルギー量、耐雑音性についてSTRAIGHTと比較し、TANDEMが優れていることを示す。基本周波数に推定誤差が含まれる場合についても検討し、数%の誤差ならば実用上問題無い性能が得られることを示す。さらに、音声分析に適した窓関数について考察する。

第3章では、人間の口から放射される音が任意の位置へ伝達する特性を推定する方法の提案を行う。本論文では、位相までを含んだ伝達関数を高SNRで推定する方法を提案する。文献⁵⁶⁾の方法により、口元と観測点に設置したマイクロフォンをの各信号から、クロススペクトル法により伝達関数が推定できることを示す。しかし、音声のエネルギーは特に高域において少なく、実環境において推定される伝達関数は、高域においてSNRが低下していることを示す。本論文では、高域のSNRを改善し、同一条件で推定した伝達関数を安定させるための後処理を提案する。

無響室で多数回収録した音声に基づいて音声と暗騒音のパワースペクトルを比較し、帯域ごとのSNRを示す。収録した多数の音声に基づき推定した多数の伝達関数から周波数毎の標準偏差を求め、高域における変動が大きいことを明らかにする。提案する後処理により、全帯域における変動が低減できることを示す。

第4章では、第3章で提案したクロススペクトル法に基づく伝達関数推定についてさらに細かく検討する。第3章で提案した推定法は、任意の窓関数を用いて波形を切り出す必要がある。切り出しに用いる窓関数の種類に応じて、得られるインパルス応答の振幅が変形する誤差が含まれることを示す。この章では、音韻・個人性により大きなばらつきを持つ音声を入力信号とした場合において、振幅変形による誤差を低減できる窓関数の条件を明らかにする。

ホワイトノイズと音声における特徴の違いである、パワースペクトルの有色性、周期性、スペクトル傾斜についてそれぞれシミュレーションを行う。窓関数による振幅変形の

影響に違いが生じる原因は、狭い帯域におけるパワースペクトルの変動であることを示す。計算機シミュレーションの結果に基づき、音声を入力信号とした場合のパワースペクトル変動において振幅変形の影響を受けにくい窓関数の条件を示す。

第5章では、音響システムの音がどのように放射されているか調べるための測定用信号を提案する。第3章で提案した推定法は、口元で収録した信号と観測点のクロススペクトルに基づいて計算しているため、音源位置を口元に設置したマイクロフォンとしている。口元に設定したマイクロフォンに基づいて推定した伝達関数の妥当性については、ダミーヘッドの口内からインパルス応答測定用の信号を用いて測定した伝達関数と比較し検討する必要がある。従来法⁵⁶⁾では、音声が大きなパワーを有する帯域において1 dB以下の誤差となることが示されている。しかし、高域ではSNRが低く信頼できる分析が行えないことが指摘されていた。そのため、第3章で示した後処理だけではなく、スピーカのインパルス応答測定においても、全帯域において高いSNRで測定可能な測定用信号が必要となる。

本論文では、この問題を解決するため、実環境の暗騒音の特性に対応して変化させ、高いSNRで測定が行える新たな測定用信号の提案を行う。この測定用信号を実環境で評価することで、環境に依存せず従来法より高い精度で測定が行えることを示す。

最後に、本論文のまとめと今後の課題を示す。

第 2 章 分析時刻に依存しない周期信号 のパワースペクトル推定法

2.1. はじめに

高品質の Vocoder を作成するためには、収録された音声から、合成に必要な基本周波数とスペクトル包絡を高精度に推定する必要がある。一般的に音声のような短時間で特性が変化する信号を分析する場合、短時間フーリエ変換^{13, 35)} が用いられる。分析対象となる信号が 2 次の確率過程である場合、短時間フーリエ変換により求めた短時間パワースペクトルが分散する。分析時刻をシフトして得た多数の短時間スペクトルを平均することにより、分散を抑圧することができる⁷⁾。

母音のように周期的に繰り返す信号では、周期性の存在がパワースペクトルの推定精度に大きく影響する。この影響を避ける方法としてピッチ同期分析⁵⁷⁾ が提案されているが、分析時刻に依存せず安定したパワースペクトルを与えるものとはならない。

音声分析変換合成法 STRAIGHT では、相補的時間窓と呼ばれる特殊な窓関数を用いることでこの問題を解決している。STRAIGHT では、窓関数と、その窓関数に基本周期の 2 倍の周期の正弦波を掛けて作成した窓関数のそれぞれから求められたパワースペクトルを適切な重みで合成することにより、この位置依存性を取り除いている。しかし、そこで用いられている窓関数や定数を設定するための根拠が明確ではなく見通しが悪い。また、正弦波による変調には、周波数分解能を低下させるという問題がある。本論文ではこの問題を解決するため、STRAIGHT よりも分析時刻への依存性を低減でき、周波数分解能の低下も生じない新たな音声分析法である TANDEM (Temporally Aligned, Non-Dispersive Envelope Measurement)⁵⁸⁾ を提案する。

本章では、周期信号を分析する場合における問題点について説明し、その問題を解決する TANDEM の原理について説明する。次に、計算機シミュレーションにより TANDEM を評価し、STRAIGHT よりも音声分析に適していることを明らかにする。

2.2. 周期信号を時間周波数分析した場合に起こる問題

ここではパルス列の時間周波数分析を例にして、周期信号を時間周波数分析した場合に起こる問題について説明する。

2.2.1 周期信号のスペクトル

まず、周期 T_0 であるパルス列とそのスペクトルを示す。

$$x(t) = \sum_{n=-\infty}^{\infty} \delta(t - nT_0) \quad (2.1)$$

$$X(\omega) = \sum_{n=-\infty}^{\infty} \delta(\omega - n\omega_0) \quad (2.2)$$

$$\omega_0 = \frac{2\pi}{T_0} \quad (2.3)$$

このように、周期 T_0 のパルス列のスペクトルは、周波数軸上で周期 ω_0 のパルス列となる。詳しい導出は、付録にて行う。Vocoderでは、声帯振動のスペクトル情報は、スペクトル包絡側に含まれる。そのため、音源は時間軸で基本周期間隔で配置したパルス列となる。すなわち、有声音は、声帯振動間隔で発生するパルス列に、スペクトル包絡を逆フーリエ変換して得られるインパルス応答を畳み込んだものとして表現される。声道形状が一定であり、スペクトル包絡 $P(\omega)$ を逆フーリエ変換して得られるインパルス応答が $p(t)$ として与えられる場合、音声は次のようにパルス列とインパルス応答との畳み込みとなる。

$$y(t) = x(t) * p(t) \quad (2.4)$$

$$\begin{aligned} Y(\omega) &= X(\omega)P(\omega) \\ &= P(\omega) \sum_{n=-\infty}^{\infty} \delta(\omega - n\omega_0) \end{aligned} \quad (2.5)$$

ここで $*$ は畳み込みを表す。 $X(\omega)$ が周波数軸上で周期 ω_0 パルス列であることから、 $Y(\omega)$ は周波数軸上で ω_0 の整数倍でのみ任意の振幅と位相を持つスペクトルとなる。つまり、任意のインパルス応答を周期的に配置することは、図 2.1 のように周波数軸上で離散化を行っていることに相当する。

音声スペクトルからスペクトル包絡を推定する問題は、離散化されたスペクトルから、元のスペクトルを復元する問題と考えることが出来る。次の節では、周期信号を任意の窓関数で切り出すことがスペクトルにどのような影響を与えるか明らかにする。

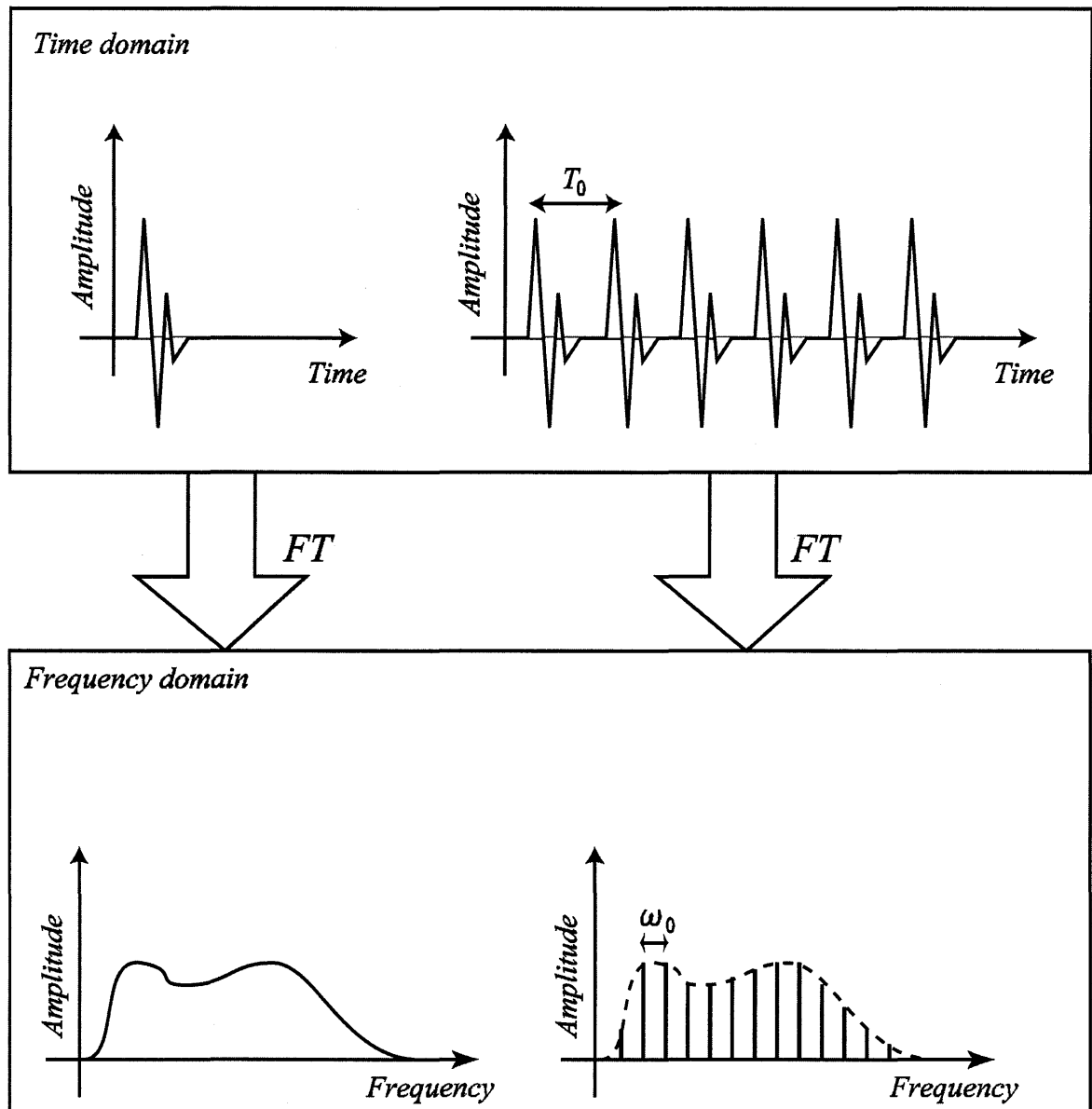


図 2.1 周期信号の周波数特性. 任意のインパルス応答 (左上) を時間軸で周期的に配置した信号 (右上) のスペクトルは, 元のインパルス応答のスペクトル (左下) を周波数軸上で離散化したもの (右下) となる.

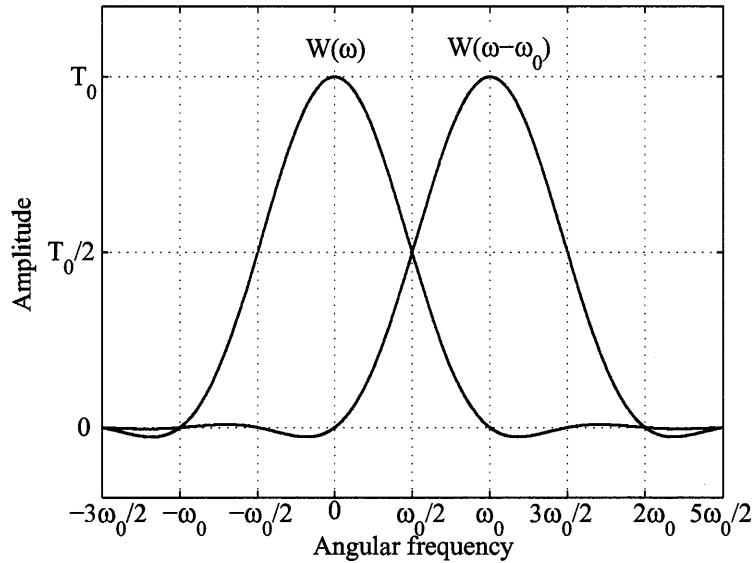


図 2.2 窓関数長を周期の 2 倍に設定した Hanning 窓のスペクトル.

2.2.2 窓関数による切り出しがスペクトルに与える影響

周期信号から得られるスペクトルは、基本周波数の整数倍において任意の振幅と位相を持つパルス列となることを示した。この節では、周期信号の短時間スペクトルを得る際に生ずる問題点について説明する。まず、周期信号の短時間スペクトルを得るために、信号を任意の窓関数により切り出す必要がある。このとき、窓関数は任意のスペクトル $W(\omega)$ を持っているものとする。周期信号は任意の時刻で切り出す必要があるため、窓関数を任意の分析時刻 τ にシフトして分析を行う。このときの窓関数のスペクトルは以下のようになる。

$$W(\omega, \tau) = W(\omega)e^{-j\omega\tau} \quad (2.6)$$

周期信号を分析時刻 τ で切り出すことは、周波数軸上で、 $W(\omega, \tau)$ と畳み込みを行うことに相当する。すなわち、図 2.2 のように、窓関数のスペクトルを ω_0 の整数倍の周波数へシフトして加算したスペクトルが得られる。

ピッチ同期分析では、窓関数のスペクトルに生ずる周期的なゼロ点が ω_0 の整数倍となるため、 ω_0 の整数倍では、分析時刻に依存せず安定したパワースペクトルを与える。しかし、隣り合う調波間では、2つのスペクトルに干渉が生じることが分かる。2つのスペクトルの相対位相は分析時刻に依存して異なるため、ピッチ同期分析を行う場合におい

ても、 ω_0 の整数倍以外の周波数では、分析時刻に依存したパワースペクトルが得られることとなる。図 2.3 は、窓関数の 2 つのスペクトルが同位相となる場合を示しており、図 2.4 は、窓関数の 2 つのスペクトルが逆位相となる場合を示している。

この影響を時間周波数領域で観測した場合、図 2.5 に示すように時間・周波数方向で周期的な谷として現れる。図 2.5 は、周期 10 ms(基本周波数 100 Hz) のパルス列を、周期の倍である 20 ms の Hanning 窓を用いてシフト時刻 1 ms で時間周波数分析を行ったものである。周期信号の時間周波数分析では、時間周波数方向に生ずる周期的な谷を取り除くことが重要な問題となる。同時に、STRAIGHT における重要な特徴である、基本周波数の整数倍においてスペクトル包絡の値を保存することも求められる。

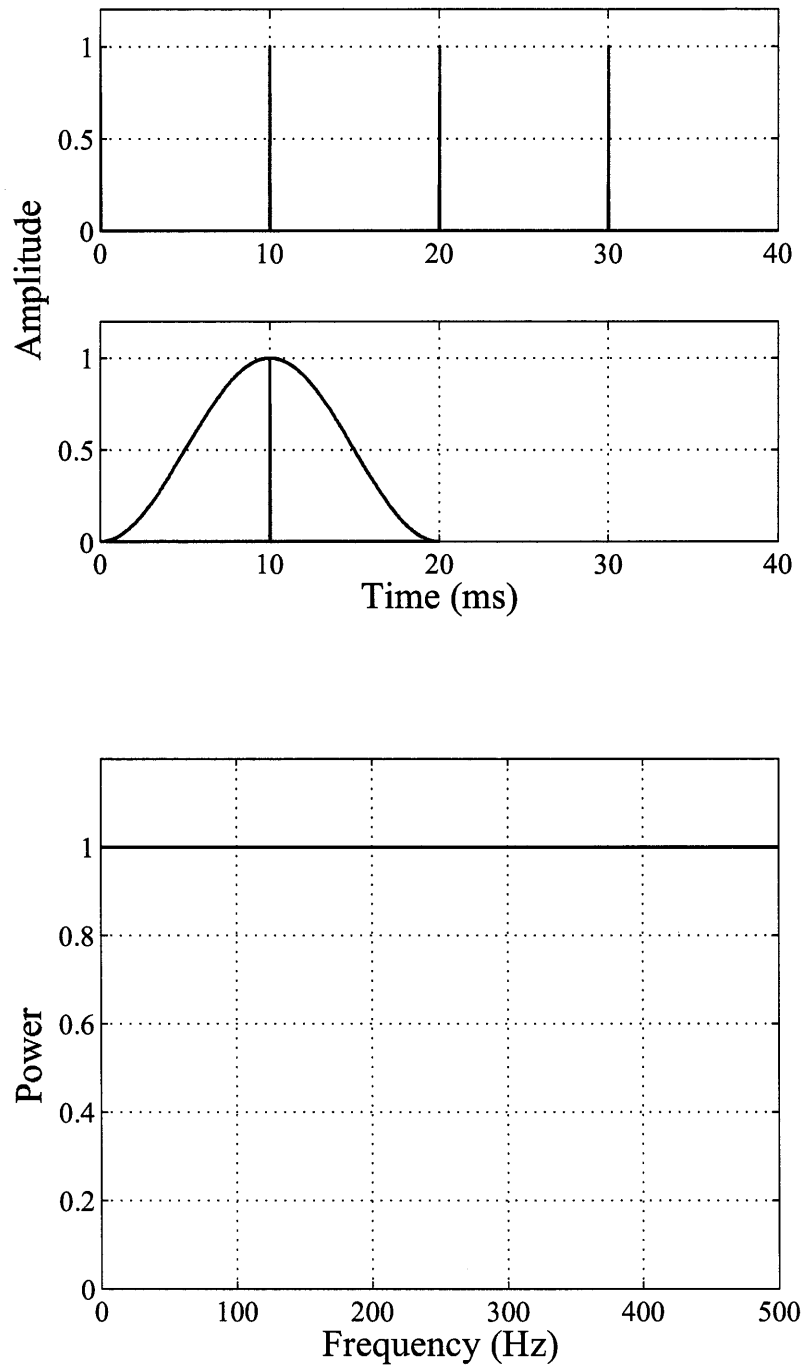


図 2.3 基本周期 10 ms のパルス列 (上) と窓関数で切り出した波形 (中), 切り出された波形のパワースペクトル (下).

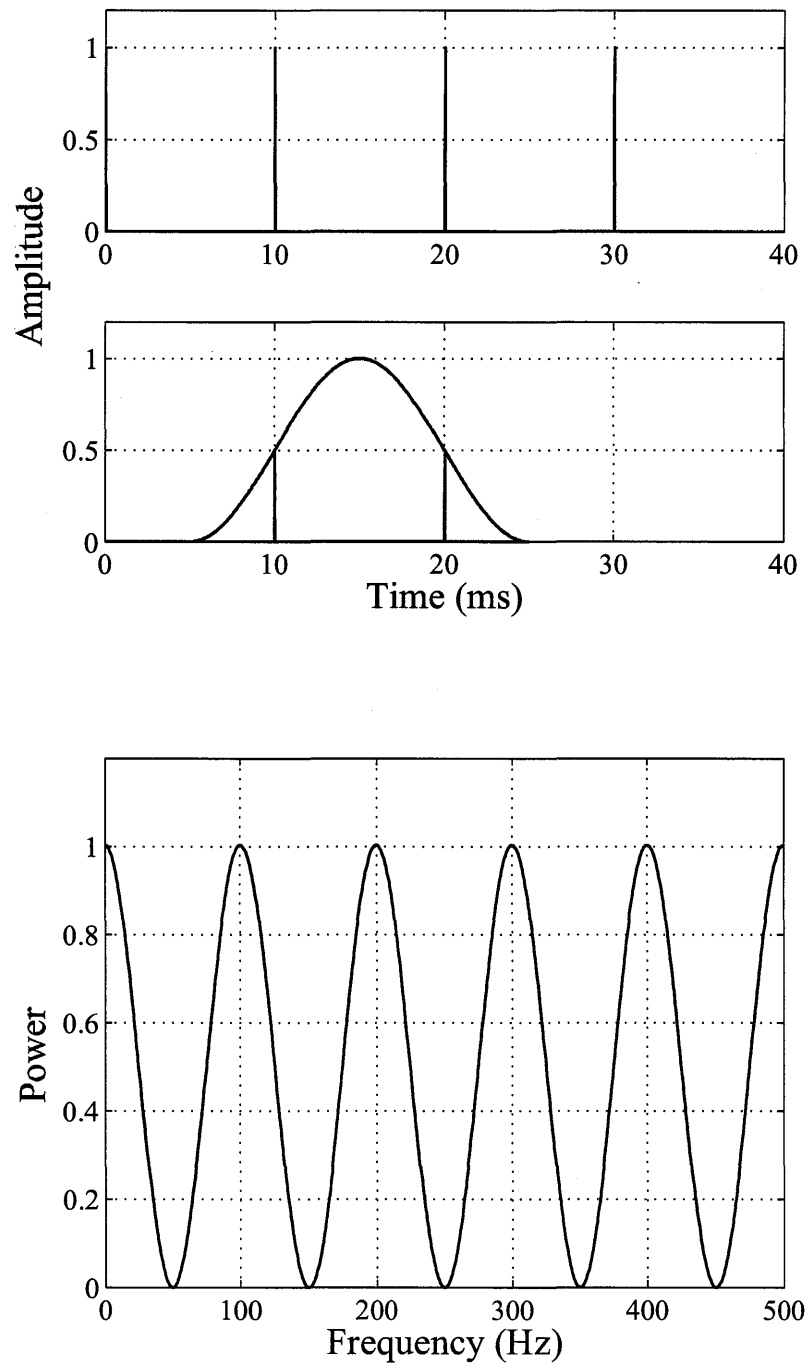


図 2.4 分析時刻を 5 ms シフトした場合の波形 (上), 窓関数で切り出された波形 (中), 切り出された波形のパワースペクトル (下).

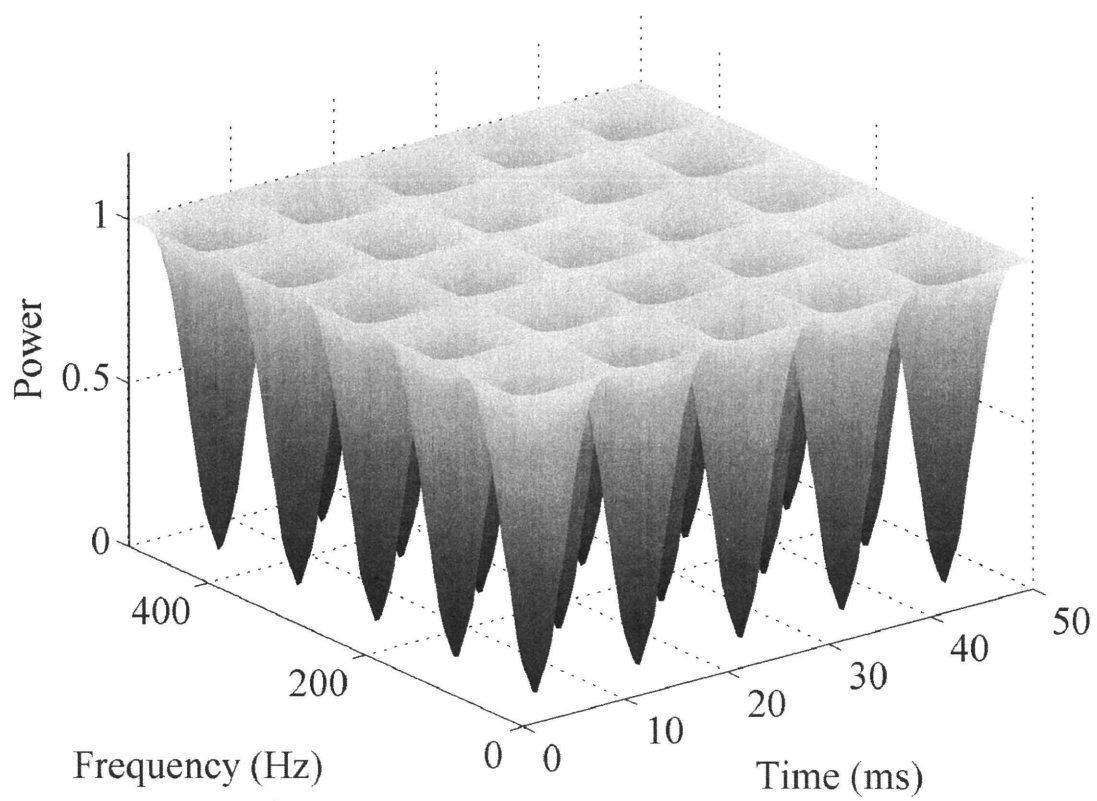


図 2.5 パルス列の時間周波数分析. 時間周波数方向に周期的な谷が認められる.

2.3. 分析時刻に依存しない周期信号のパワースペクトル推定法：TANDEM

これまでの節により，窓関数により周期信号を切り出すことで，窓関数のスペクトルが隣り合う調波で干渉することを示した．また，その影響が時間周波数領域で周期的に生じる谷として現れることを示した．この干渉による影響を簡単に解決する方法は，窓関数長を増加させることである．窓関数長の増加はメインローブを鋭くする効果がある．そのため，調波間における干渉の影響を低減でき，分析時刻に影響しないスペクトルを与える．しかしながら，母音のような周期信号では声道形状が短時間で変化していることから，長い窓関数を用いた分析は時間変動とは異なる弊害をもたらすこととなる．また，周波数方向への変動が大きく現実的な解決法とは言いがたい．STRAIGHTでは，相補的時間窓と呼ばれる特殊な窓を用いることで，この影響を低減してきた．相補的時間窓の具体的な内容については付録で説明する．

本論文では，周期信号から分析時刻に依存した成分を取り除くことが出来る新しい分析法 TANDEM(Temporally Aligned, Non-Dispersive Envelope Measurement) を提案する．TANDEMでは，1つの窓関数を異なる時間に配置し，得られたパワースペクトルを合成する方法である．TANDEMに用いる窓関数はTANDEM窓と定義する．TANDEM窓とは図2.6上段に示されるように，ある時刻 τ と， τ から周期 T_0 の半分シフトした2つの窓関数のことである．以下では，1つの窓関数で切り出した信号のパワースペクトルにおける分析時刻依存性を明らかにし，TANDEMを用いることで，この分析時刻への依存性が実質的に除去できることを示す．

2.3.1 1つの窓関数による分析時刻依存性

TANDEM窓 $w(t)$ と分析対象となる周期信号 $x(t)$ に，以下のような仮定を設定する．

1. 周期信号 $x(t)$ の基本周期は $T_0 = 1/f_0$ であり，分析前より既知である．
2. TANDEM窓 $w(t)$ は，有限の長さを持つ．
3. TANDEM窓 $w(t)$ のスペクトル $W(\omega)$ は，基本周波数 f_0 Hz 以上において振幅が0である．

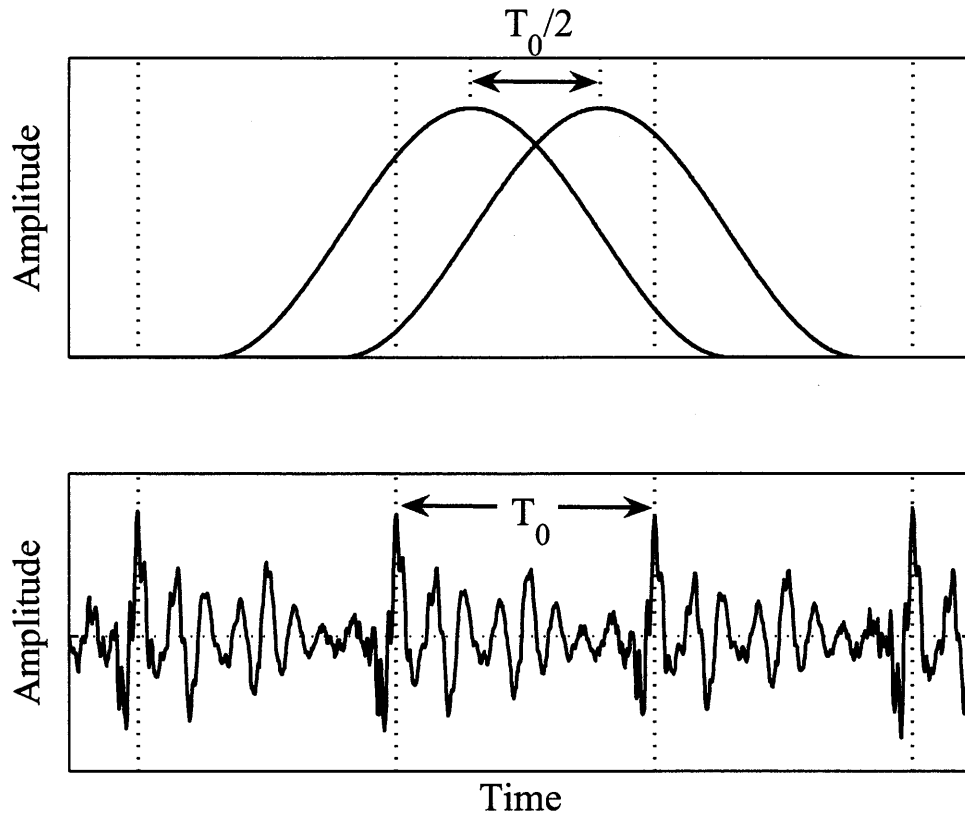


図 2.6 TANDEM 窓 (上段) と周期 T_0 の周期信号の例 (下段). TANDEM 窓とは, 任意の窓関数を周期の半分だけシフトした 2 つの窓関数のことである.

任意の時刻 τ において, TANDEM 窓 $w(t)$ により切り出した信号の波形を $s(t, \tau)$ とすると, $s(t, \tau)$ は以下の式により表される

$$s(t, \tau) = x(t)w(t - \tau) \quad (2.7)$$

この窓と対象となる信号の積である $s(t, \tau)$ のスペクトルは, 周波数領域においてそれぞれのスペクトルの畳込みとなる. ここで, 窓関数のスペクトル $W(\omega)$ において定めた 3 番目の仮定より, ある周波数 ω における $s(t, \tau)$ のスペクトル $S(\omega, \tau)$ は, ω に隣接する 2 つの調波成分からの影響だけを考慮すれば良い. 周波数軸上で平行移動しても同じ議論が成立するため, ここでは隣り合う 2 つの調波についての振舞いを調べる. すなわち無限に連なるパルス列であった $X(\omega)$ を以下のように 2 つのパルスに限定して考える.

$$X(\omega) = \delta(\omega) + \alpha e^{j\beta} \delta(\omega - \omega_0) \quad (2.8)$$

振幅・位相は隣り合う調波の相対的な値が重要となるため、片側の調波が相対的に異なる値を持っているとして議論する。ここで、 α, β は、実数であり、 $\omega_0 = 2\pi f_0$ である。

この振幅・位相が相対的に異なる2つのパルスに対して、分析時刻 τ における窓関数のスペクトル $W(\omega, \tau) = W(\omega) \exp(-j\omega\tau)$ を畳み込めば、分析時刻に依存したスペクトル $S(\omega, \tau)$ が求まる。

$$\begin{aligned} S(\omega, \tau) &= X(\omega) * W(\omega, \tau) \\ &= e^{-j\omega\tau} \left(W(\omega) + \alpha e^{j(\tau\omega_0 + \beta)} W(\omega - \omega_0) \right) \end{aligned} \quad (2.9)$$

このパワースペクトルは以下となる。

$$\begin{aligned} |S(\omega, \tau)|^2 &= W^2(\omega) + \alpha^2 W^2(\omega - \omega_0) \\ &\quad + 2\alpha W(\omega) W(\omega - \omega_0) \cos(\omega_0\tau + \beta) \end{aligned} \quad (2.10)$$

この第3項が、窓の時刻 τ に応じて正弦波状に変動する成分を表している。これが問題としている、窓関数と分析される周期信号の相対位置(分析時刻 τ)に起因する干渉の影響である。

2.3.2 2つ目の窓関数による干渉の抑圧

時刻を基本周波数の半分だけ移動させた TANDEM 窓 $w(t - (\tau + T_0/2))$ により信号を切り出す。この信号のスペクトルは以下になる。

$$\begin{aligned} S(\omega, \tau + T_0/2) &= X(\omega) * W(\omega, \tau + T_0/2) \\ &= e^{-j\omega(\tau + T_0/2)} \left(W(\omega) + \alpha e^{j((\tau + T_0/2)\omega_0 + \beta)} W(\omega - \omega_0) \right) \end{aligned} \quad (2.11)$$

ここで、右辺の $e^{j((\tau + T_0/2)\omega_0 + \beta)}$ を以下のように展開する。

$$e^{j((\tau + T_0/2)\omega_0 + \beta)} = e^{j(\tau\omega_0 + \beta)} e^{j\omega_0 T_0/2} \quad (2.12)$$

$\omega_0 = 2\pi/T_0$ を代入すると $e^{j\omega_0 T_0/2}$ は以下となる。

$$e^{j\omega_0 T_0/2} = e^{j\pi} = -1 \quad (2.13)$$

式2.13, 2.12 より、式2.11 は以下のようになる。

$$S(\omega, \tau + T_0/2) = e^{-j\omega(\tau + T_0/2)} \left(W(\omega) - \alpha e^{j(\tau\omega_0 + \beta)} W(\omega - \omega_0) \right) \quad (2.14)$$

したがって、パワースペクトル $|S(\omega, \tau + T_0/2)|^2$ は、次のようになる。

$$\begin{aligned} |S(\omega, \tau + T_0/2)|^2 &= W^2(\omega) + \alpha^2 W^2(\omega - \omega_0) \\ &\quad - 2\alpha W(\omega)W(\omega - \omega_0) \cos(\omega_0\tau + \beta) \end{aligned} \quad (2.15)$$

両者の平均を求めると第3項が相殺して消え、窓の時刻 τ に依存しない項のみが残る。

$$\frac{|S(\omega, \tau + T_0/2)|^2 + |S(\omega, \tau)|^2}{2} = W^2(\omega) + \alpha^2 W^2(\omega - \omega_0) \quad (2.16)$$

すなわち、TANDEM窓により得られたパワースペクトルを平均することで、隣り合う調波間隔において分析時刻 τ に依存しないパワースペクトルが得られる。

しかし、現実には存在する有限長の窓関数は、窓関数について設定した仮定を満足しない。式2.16は、窓関数のスペクトルが0 Hzから f_0 Hzまでは分析時刻に依存しないことを証明している。そのため、TANDEM窓では、基本周波数より高い周波数帯域におけるエネルギーが少ない、サイドローブの小さな窓を必要とする。音声分析では窓関数長が短いことも要求されるため、これらを満たす窓関数を計算機シミュレーションによる明らかにする。また、分析信号における周期 T_0 が既知であるという仮定を満足しない(例えば基本周波数推定において誤差が含まれる)場合についても、分析結果にどのような影響を与えるか検討する。

2.3.3 TANDEMによる時間周波数分析

基本周期の2倍の窓関数長を持つHanning窓を用いたパルス列の時間周波数分析例を示す。TANDEMを用いた時間周波数分析では、図2.7のように、分析時刻に依存して生ずる干渉の影響を取り除き、時間方向で周期的に現れていた谷を除去する効果がある。ただし、本来平坦であるはずのパルスのパワースペクトルが周波数方向に振動していることも分かる。

この周波数方向の変動は、 ω_0 を周期とした正弦波的な振動として現れる。このような周期信号の変動は、 ω_0 の長さを持つ矩形窓を畳み込むことで除去できる。矩形窓を畳み込むことにより、図2.7で認められた周波数軸上での周期的な振動成分が取り除かれ図2.8のようなスペクトログラムが得られる。

TANDEMを音声分析合成システムSTRAIGHTで用いられているSTRAIGHTスペクトルに置き換えるためには、音声合成系で生ずる影響までを考慮して、パワースペクトルを周波数軸上で補償する必要がある。文献⁵⁹⁾では、TANDEMをSTRAIGHTにおけ

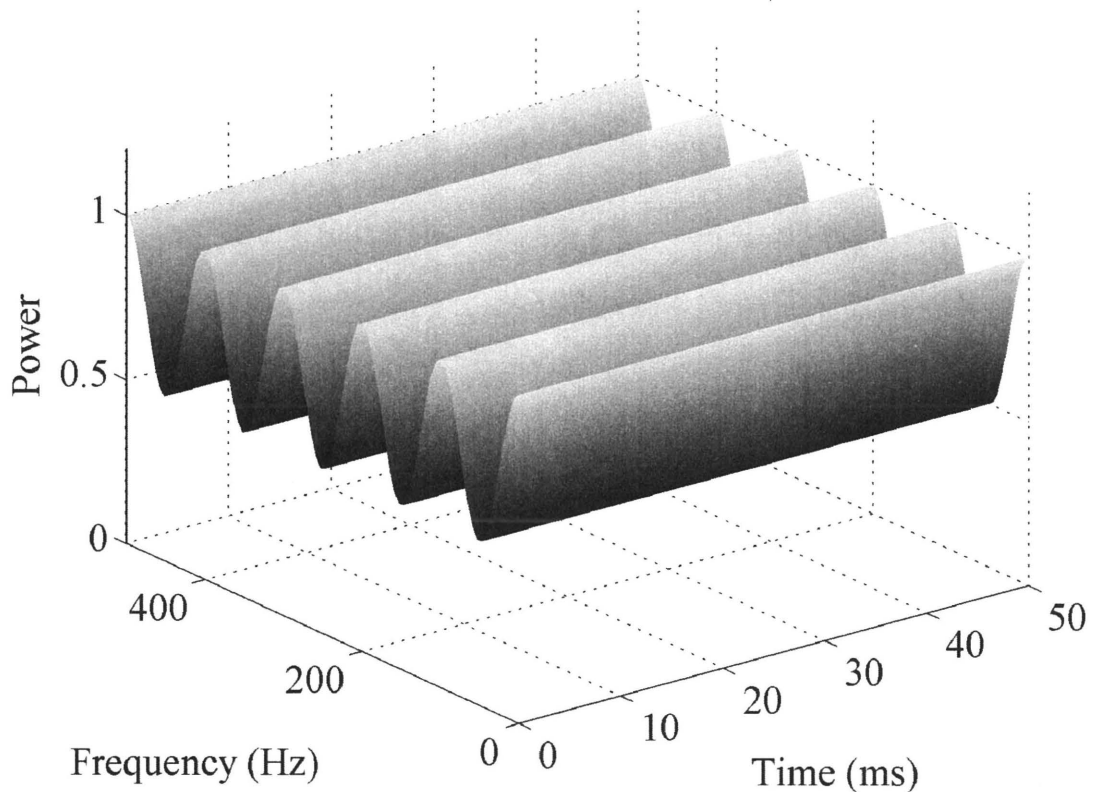


図 2.7 TANDEM を用いたパルス列の時間周波数分析例. TANDEM は時間方向で周期的に現れた谷が取り除き, 分析時刻によらず安定したパワースペクトルを与える. TANDEM におけるパワースペクトルの周波数変動は ω_0 を周期とする振動として現れる.

る相補的時間窓の代わりに使用するための平滑化と, 基本周波数の整数倍においてスペクトル包絡と同等のパワーを持つように補償する後処理を提案している. これらの平滑化処理後のスペクトル包絡は, STRAIGHT スペクトルとほぼ同等となる. TANDEM では, 処理時間も現状の STRAIGHT より早い利点がある.

2.4. 評価

この節では, 2.3.1 節で設定した仮説を近似的に満たし, 音声分析に適した TANDEM 窓の選定を行う. TANDEM 窓に適する条件である, サイドローブが小さい窓関数¹⁶⁾を用いて TANDEM を評価する. また, TANDEM を用いない短時間フーリエ変換や STRAIGHT

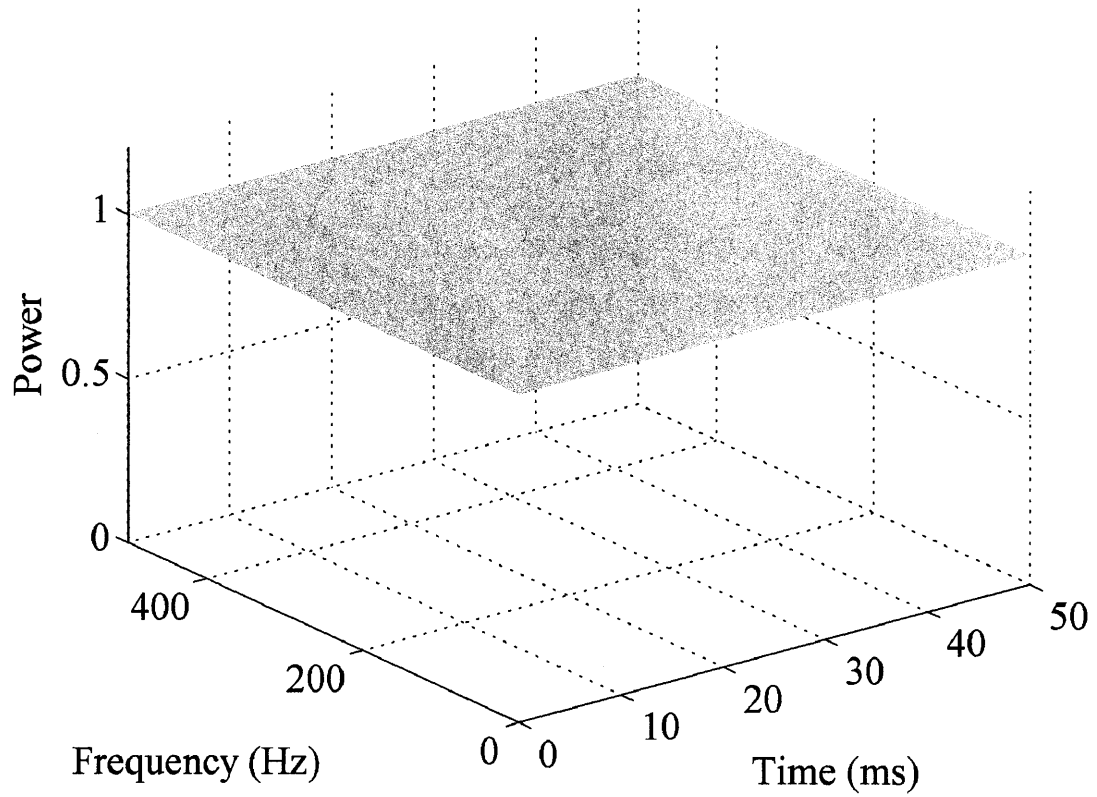


図 2.8 TANDEM により得られたパワースペクトルを周波数方向に平滑化して得られたスペクトログラム．時間周波数方向に滑らかな応答が得られる．

における相補的時間窓と比較し，TANDEM の有効性を示す．

2.4.1 TANDEM に適した窓関数の候補

本論文では，サイドローブが小さく有限の長さで定義される窓関数として，Hanning 窓，Blackman 窓，Nuttall 窓¹⁷⁾ を用いた．

これらの窓関数は以下の式により表される．

$$w_{hanning}(t) = 0.5 - 0.5 \cos(2\pi t) \quad (2.17)$$

$$w_{blackman}(t) = 0.42 - 0.5 \cos(2\pi t) + 0.08 \cos(4\pi t) \quad (2.18)$$

$$\begin{aligned} w_{nuttall}(t) &= 0.355768 - 0.487396 \cos(2\pi t) \\ &+ 0.144232 \cos(4\pi t) - 0.012604 \cos(6\pi t) \end{aligned} \quad (2.19)$$

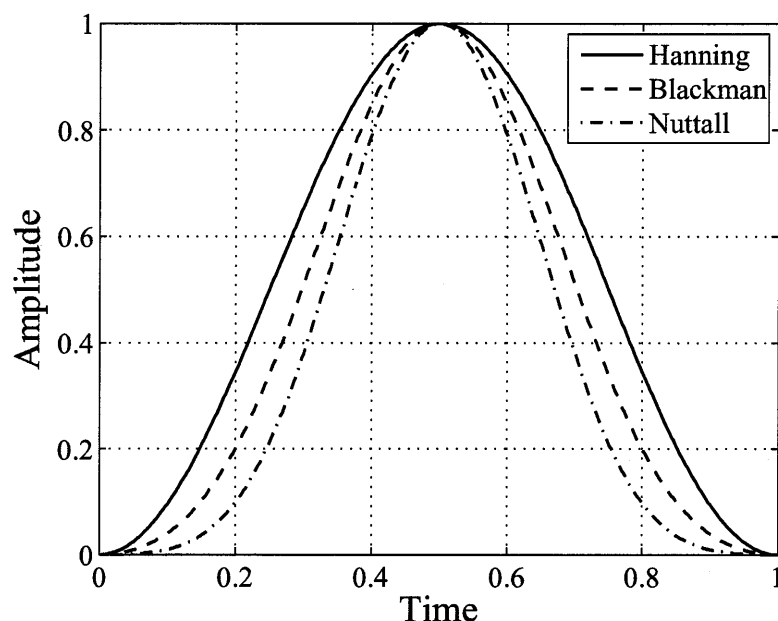


図 2.9 3 種類の窓関数の時間波形.

全ての窓関数は、 $0 \leq t \leq 1$ 以外での振幅は 0 である．これらの窓関数の時間波形を 2.9 に、パワースペクトルを図 2.10 に示す．

3 種類の窓関数の第 1 サイドローブは Hanning 窓でおよそ -30 dB, Blackman 窓でおよそ -60 dB, Nuttall 窓ではおよそ -90 dB となる．ただし、メインローブは Hanning 窓が最も鋭く、サイドローブの最も小さい Nuttall 窓はメインローブが最も広がっている．音声分析では、高い時間分解能を要求するため、TANDEM の性能と時間分解能との関係を明らかにする必要がある．次節では、計算機シミュレーションにより TANDEM の性能評価を行う．

2.4.2 TANDEM 窓による性能の違い

はじめに、計算機シミュレーションを用いて窓関数の持続時間と TANDEM の性能との関連を明らかにする．計算機シミュレーションは、同一の 10 ms のホワイトノイズを時間的に繰り返して並べた信号を用いた．これは周期信号ではあるが、パワーが無相関であるため、評価用信号として有効である．以下の実験ではこのホワイトノイズを 100 回無作為に生成して用いた．その他の実験条件は表 2.1 にまとめる．

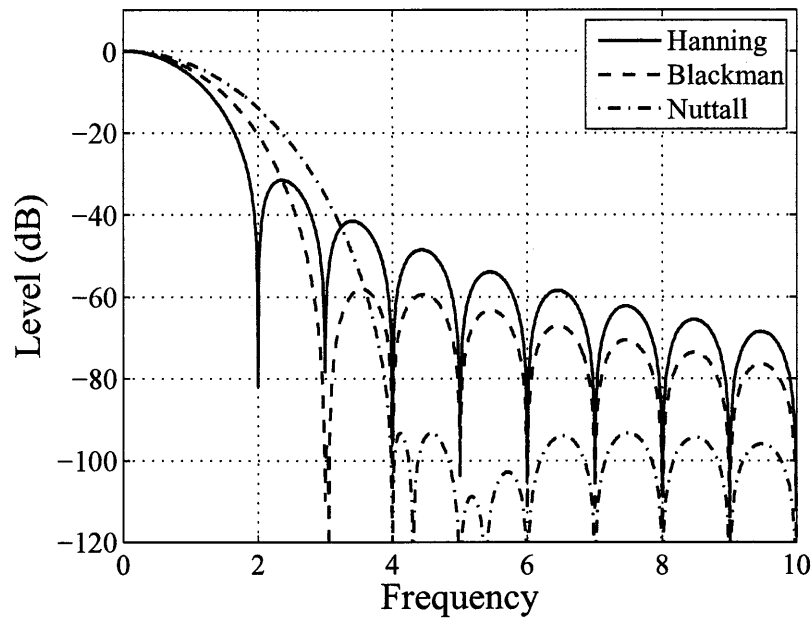


図 2.10 3 種類の窓関数のパワースペクトル.

表 2.1 実験条件

周期信号の種類	時間軸上で周期的に繰り返したホワイトノイズ
周期	10 ms
サンプリング周波数	48 kHz
使用した窓関数	Hanning 窓, Blackman 窓, Nuttall 窓
比較対象	単一窓の短時間フーリエ変換, STRAIGHT の相補的時間窓
試行回数	100 回
窓関数長	周期の 2 倍から 4 倍まで

TANDEM 窓の性能は, 分析時刻がパワースペクトルに与える影響が小さいほど優れているといえる. そこで, 評価基準は, 分析時刻 τ を 1 周期分変化させスペクトルの振幅の RMS 値に対する標準偏差 (以下では時間変動とする) とした. この時間変動 E は, 以下の式により求められる.

$$E = \sqrt{\frac{\int_{-\infty}^{\infty} \int_0^{T_0} (B(\omega, \tau) - \bar{B}(\omega))^2 d\tau d\omega}{T_0 \int_{-\infty}^{\infty} \bar{B}(\omega)^2 d\omega}} \quad (2.20)$$

$$\bar{B}(\omega) = \frac{1}{T_0} \int_0^{T_0} B(\omega, \tau) d\tau \quad (2.21)$$

ここで $B(\omega)$ は振幅スペクトルを表す，短時間フーリエ変換における $B(\omega)$ は切り出された波形の振幅スペクトル，TANDEM 窓では，2つの窓から得られるパワースペクトルを平均したものの平方根である．

窓関数長は全ての窓関数において，周期の2倍から4倍までとした．ただし，窓関数の長さや時間分解能との関係は窓関数の種類により異なるため，窓関数の持続時間¹⁴⁾を用いて時間分解能の評価を一般化する．窓関数の持続時間 σ_t は以下の式により与えられる．

$$\sigma_t^2 = \frac{\int_{-\infty}^{\infty} (t - \langle t \rangle)^2 |w(t)|^2 dt}{\int_{-\infty}^{\infty} |w(t)|^2 dt} \quad (2.22)$$

$$\langle t \rangle = \frac{\int_{-\infty}^{\infty} t |w(t)|^2 dt}{\int_{-\infty}^{\infty} |w(t)|^2 dt} \quad (2.23)$$

ここで $w(t)$ は窓関数の時間波形である． σ_t が小さいほど時間分解能が高く，音声分析に適しているといえる．

2.4.3 単一の窓を用いた短時間フーリエ変換との比較

はじめに，TANDEM と単一の窓を用いた短時間フーリエ変換との比較を行う．図 2.11, 2.12, 2.13 は，それぞれの窓関数を単一で用いた場合，TANDEM 窓として用いた場合の結果を比較している．図 2.14 は，単一の窓関数から求めた時間変動を3種類の窓について重ねて表示している．TANDEM 窓における持続時間は，2つの窓関数の自乗平均値から計算した．

単一の窓を用いた短時間フーリエ変換についての結果(図 2.14)を窓関数毎に比較すると，窓関数の種類によらず持続時間と時間変動とがほぼ一致していることが分かる．これは，窓関数の持続時間はスペクトルにおけるメインローブの広さに関連しているため，

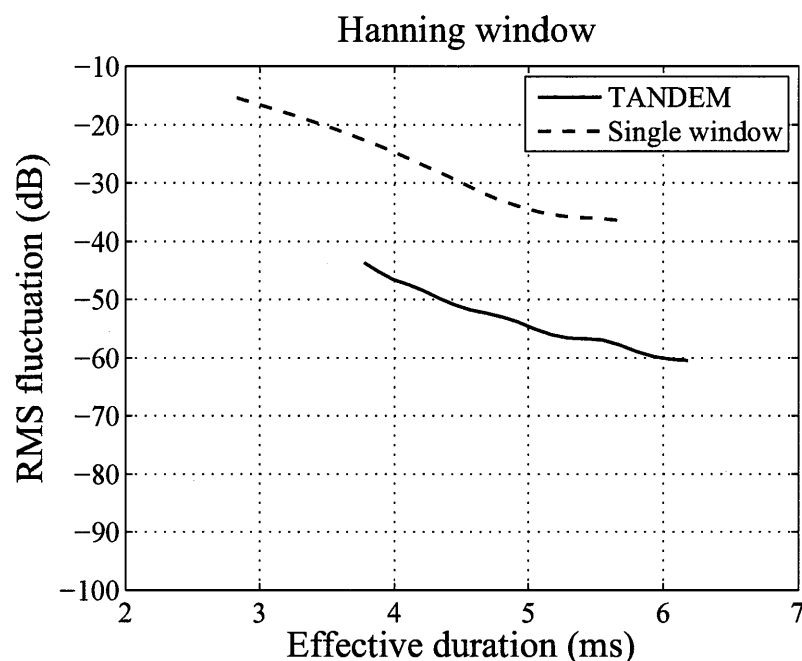


図 2.11 単一の窓関数を用いた短時間フーリエ変換を用いた場合との比較. Hanning 窓の場合.

同一の持続時間における窓関数の周波数分解能が等しく、同一の周波数分解能で生じる干渉のエネルギーが等しいことを示している.

窓関数の持続時間を比較すると、TANDEM 窓では単一の窓関数と比べると長い. これは、TANDEM は時間軸でシフトした 2 つの窓関数を用いているためである. しかし、全ての窓関数において同一の持続時間における時間変動を比較すると、Hanning 窓で約 20 dB, Blackman 窓で約 35 dB, Nuttall 窓では約 50 dB 時間変動を低減できる.

窓関数長の変化に伴う時間変動の低減については、窓関数の種類により顕著な差が認められる. まず、Hanning 窓では周期の 2 倍から 4 倍にかけて、若干の変動があるものの単調減少となる. 一方 Blackman 窓と Nuttall 窓では、窓関数長の増加に伴う時間変動の減少が単調ではなく、ある時間変動まで単調減少となり、それ以上の増加に対し大きな改善が認められない.

これは、窓関数のサイドローブと、周波数分解能が原因である. 窓関数の第 1 サイドローブは Hanning 窓で約 -30 dB, Blackman 窓で約 -60 dB, Nuttall 窓で約 -90 dB である. 図 2.2 より、周期の 2 倍の窓関数長を持つ Hanning 窓は、メインローブの幅が ω_0 と一致している. 他の窓関数では図 2.10 のように、Hanning 窓よりメインローブが広い. TANDEM

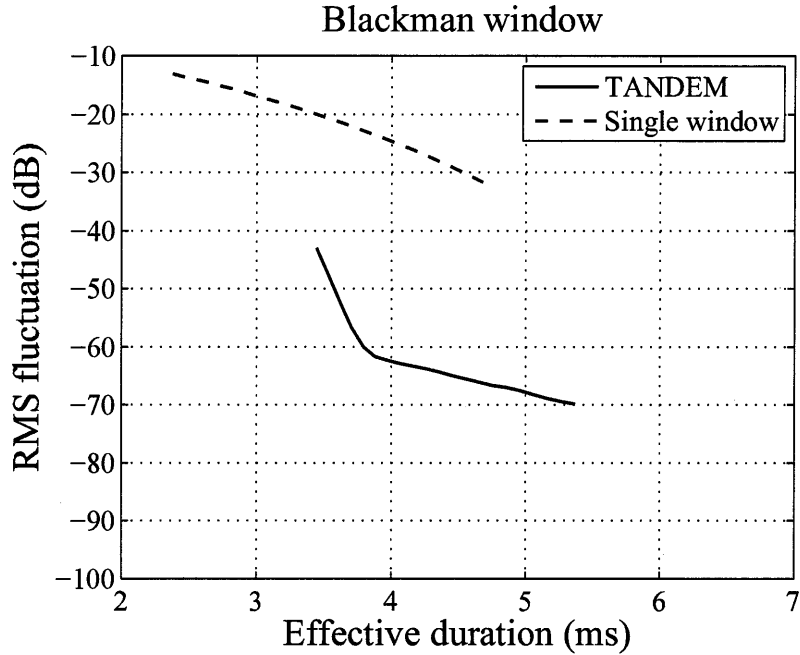


図 2.12 単一の窓関数を用いた短時間フーリエ変換を用いた場合との比較. Blackman 窓の場合.

では、隣り合う調波間についてのみ時間変動が存在しないため、メインローブが調波間隔よりも広い周期の 2 倍の Blackman 窓や Nuttall 窓では調波間外に存在するメインローブのエネルギーだけ時間変動に影響を与えることとなる. すなわち、Blackman 窓と Nuttall 窓の時間変動が窓関数長の増加に伴い急峻に減衰するのは、窓関数長の増加により調波間外に存在するメインローブのエネルギーが減衰し、その結果として時間変動が減衰したと解釈出来る. 時間変動の減衰が緩やかになる持続時間では、メインローブが調波間内に実質的に含まれる、あるいは調波間外に存在するメインローブのエネルギーがサイドローブと同程度になっていると考えられる.

2.4.4 STRAIGHT における相補的時間窓との比較

次に従来法である STRAIGHT の相補的時間窓と TANDEM との比較を行う. 図 2.15 に、窓関数の持続時間 σ_t と時間変動との関係を示す. STRAIGHT の相補的時間窓の持続時間は、加算に用いる係数を乗じた 2 つの窓関数の自乗平均値から計算した. TANDEM による結果を STRAIGHT における相補的時間窓と同じ持続時間の場合について比較す

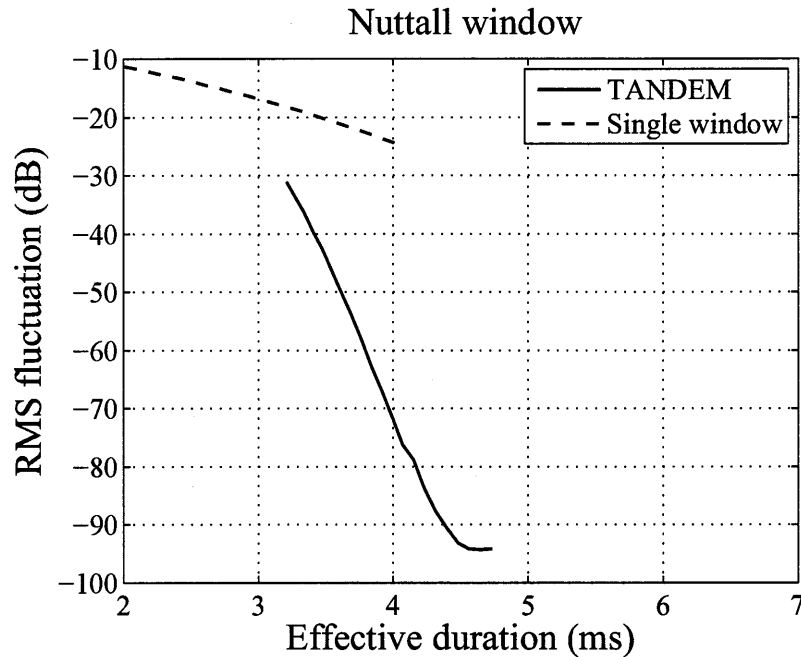


図 2.13 単一の窓関数を用いた短時間フーリエ変換を用いた場合との比較. Nuttall 窓の場合.

ると, Blackman 窓では 10 dB 程度, Nuttall 窓では 20 dB 程度時間変動が改善している.

周期信号の分析という観点では, 長い窓関数を用いた Nuttall 窓が適しているといえる. しかし, 音声分析においては時間分解能に優れた持続時間の短い窓が適している. また, 実環境における音声では声帯振動間隔は完全に一定ではなく揺らぎが含まれる. さらに, 収録環境に応じて雑音が含まれる. そこで, 周期信号に雑音が含まれている場合, 基本周波数に誤差が含まれている場合について同様の計算機シミュレーションを行うことで, 実環境で収録された音声分析に適した TANDEM 窓について論じる.

2.4.5 TANDEM の耐雑音性の評価

前節では, Blackman 窓や Nuttall 窓における TANDEM 窓が, 短時間フーリエ変換や STRAIGHT の相補的時間窓に比べ, 時間分解能・時間変動ともに優れていることを示した. この節では, 周期信号に雑音が含まれている場合について評価を行う. 計算機シミュレーションの条件は前節と同様にした. 雑音はホワイトノイズとした. SNR は 20 dB から 60 dB まで 10 dB 刻みとし, 各 SNR について時間変動を求め, 窓関数毎に比較した.

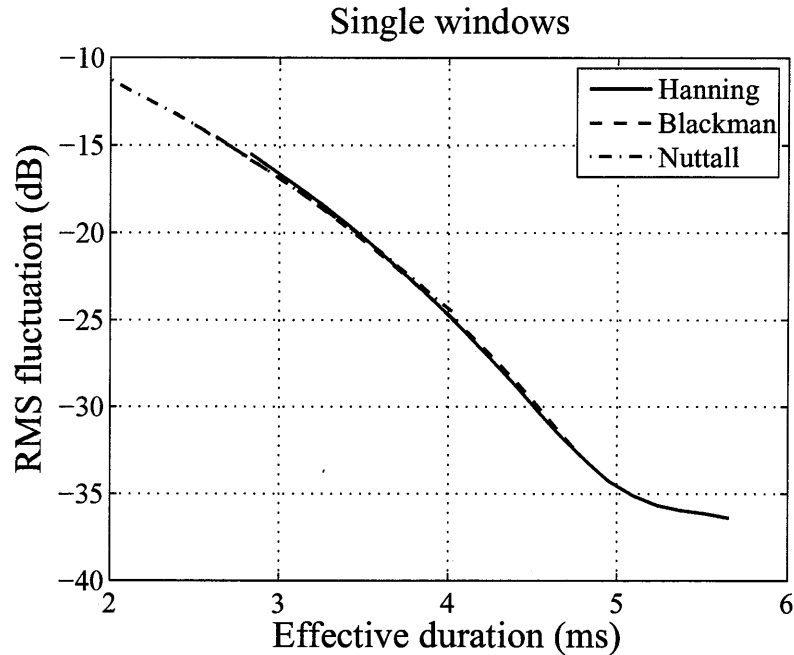


図 2.14 単一の各窓関数から求めた時間変動の比較. 単一の窓関数を用いた短時間フーリエ変換では, 窓の持続時間と時間変動とがほぼ一致する.

図 2.16, 2.17, 2.18, 2.19, 2.20 に各 SNR と時間変動との関係を示す. STRAIGHT における相補的時間窓と Hanning 窓とを比較すると, SNR が低くなるに伴い相補的時間窓における時間変動が相対的に増加していくことが分かる. SNR が 60dB, 50 dB においては, 同一の持続時間において, 相補的時間窓の時間変動が Hanning 窓を用いた TANDEM 窓の時間変動より小さい. 一方, SNR 40 dB 以下となると, Hanning 窓の時間変動のほうが小さくなる. SNR 20 dB では, 今回のシミュレーション条件の TANDEM 窓全てにおいて, 相補的時間窓より小さい時間変動となっていることが分かる. すなわち TANDEM 窓は相補的時間窓と比べ耐雑音性に優れているといえる.

周期信号に雑音が存在する場合, 全ての窓関数において, SNR より 10 dB 程度小さいレベルまで時間変動が減衰し, それ以降では減衰しないことが分かる.

2.4.6 基本周波数の誤差が時間変動に与える影響

基本周波数に含まれる誤差が時間変動にどの程度影響しているか検討する. 計算機シミュレーションの条件はこれまでと同様とし, TANDEM 窓を構成する際の基本周波数

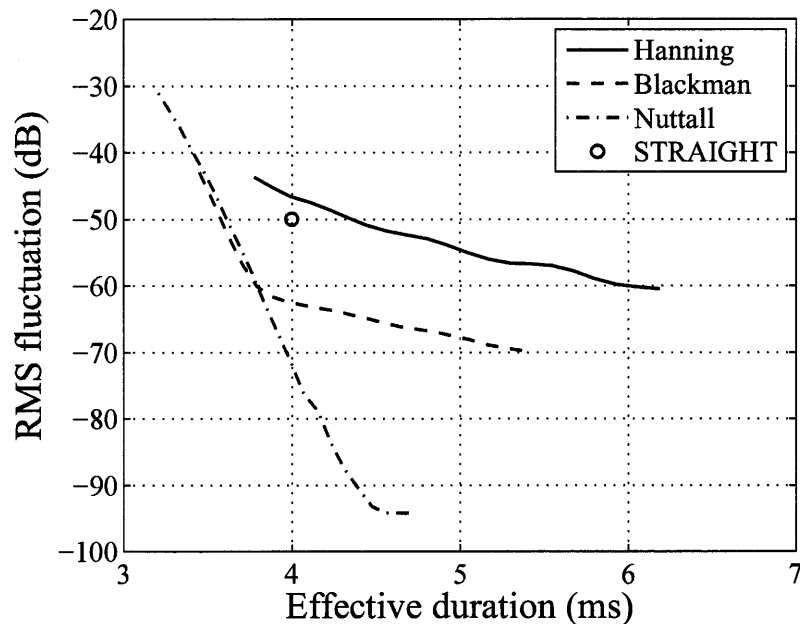


図 2.15 窓関数の持続時間と時間変動との関係. STRAIGHT と比較しても, 持続時間, 時間変動ともに優れた窓が存在する.

に誤差を加えて計算機シミュレーションを行った. 基本周波数の誤差は 1%, 3%, 10% とし, 真値より低い周波数となるようにした. すなわち, 基本周波数の真値が 100 Hz に対し 1% 誤差が生じる条件では, 基本周波数を 99 Hz として TANDEM 窓を設計することとする.

図 2.21, 2.22, 2.23 に計算機シミュレーションの結果を窓関数毎に示す. 全ての窓関数において, 基本周波数の誤差が増えるに従い窓関数の持続時間が増加している. これは, 基本周波数を真値より低い値となるよう見積もったため, 相対的に窓関数長が長くなったことが原因である.

基本周波数の誤差が時間変動に与える影響は, 窓関数毎に大きな差が存在する. Hanning 窓では, 1 % の基本周波数の誤差は, 時間変動へほとんど影響しない. 3 % 誤差が含まれる場合でも同一の持続時間ならば 5 dB 程度の性能劣化となる. 一方 Blackman 窓と Nuttall 窓では, 時間変動が 60 dB 以下の極めて小さい領域において, 基本周波数の数%の誤差が 10 dB 以上の大きな性能劣化に繋がっている.

図 2.24 は, 基本周波数が 10 % 誤差を含む場合の TANDEM 窓の時間変動について 3 種類の窓関数の結果を重ねて表示している. このように, 基本周波数に誤差が含まれる場

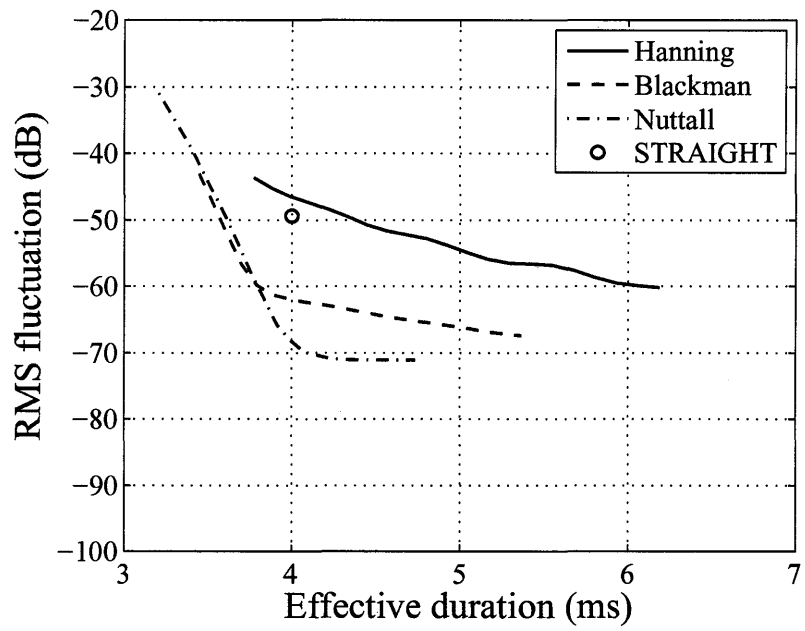


図 2.16 窓関数の種類・長さや時間変動との関係. SNR60 dB における結果.

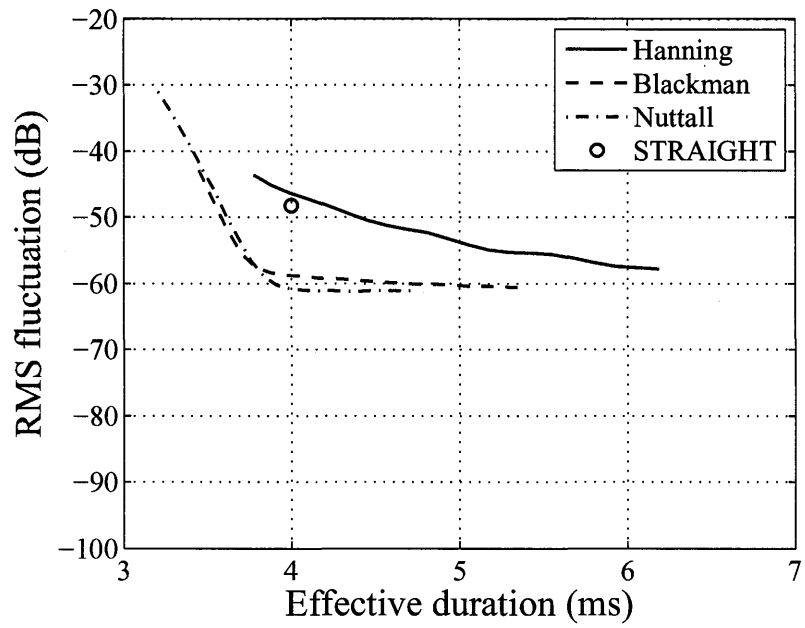


図 2.17 窓関数の種類・長さや時間変動との関係. SNR50 dB における結果.

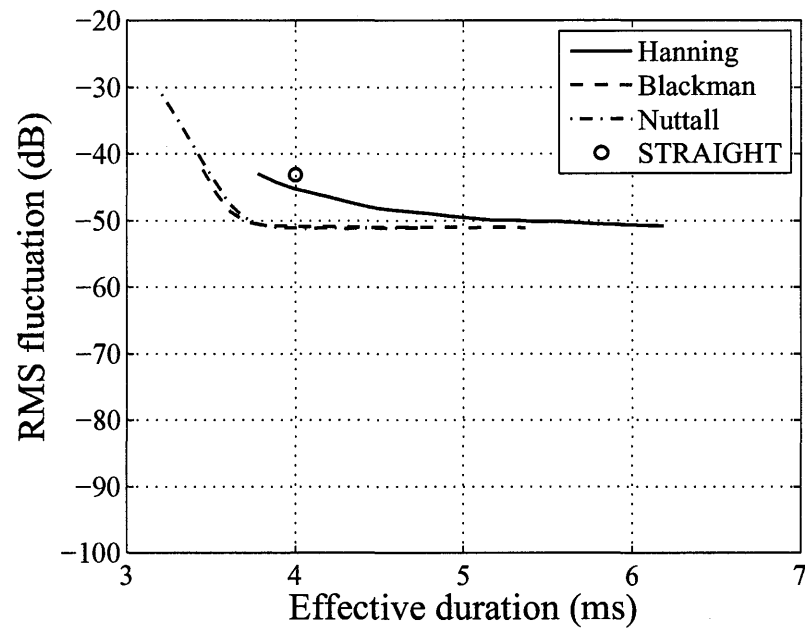


図 2.18 窓関数の種類・長さと言時間変動との関係. SNR40 dB における結果.

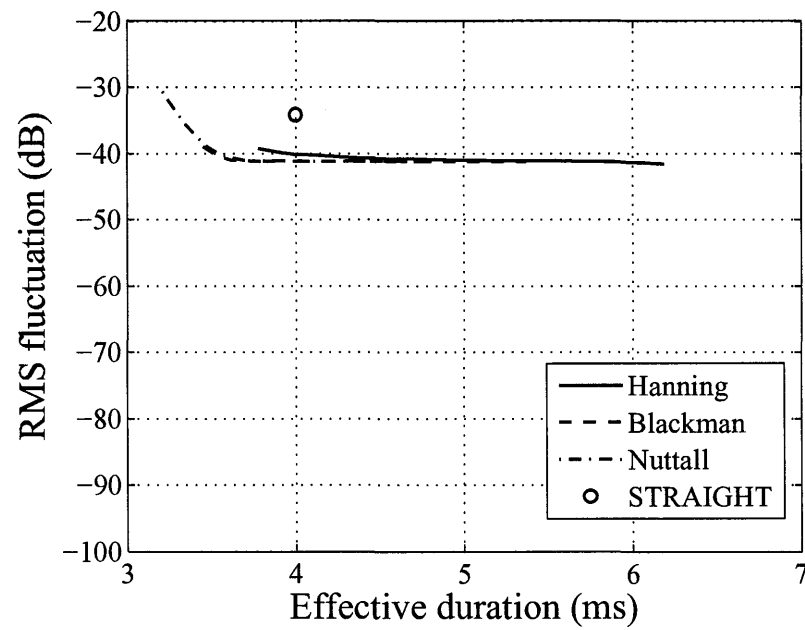


図 2.19 窓関数の種類・長さと言時間変動との関係. SNR30 dB における結果.

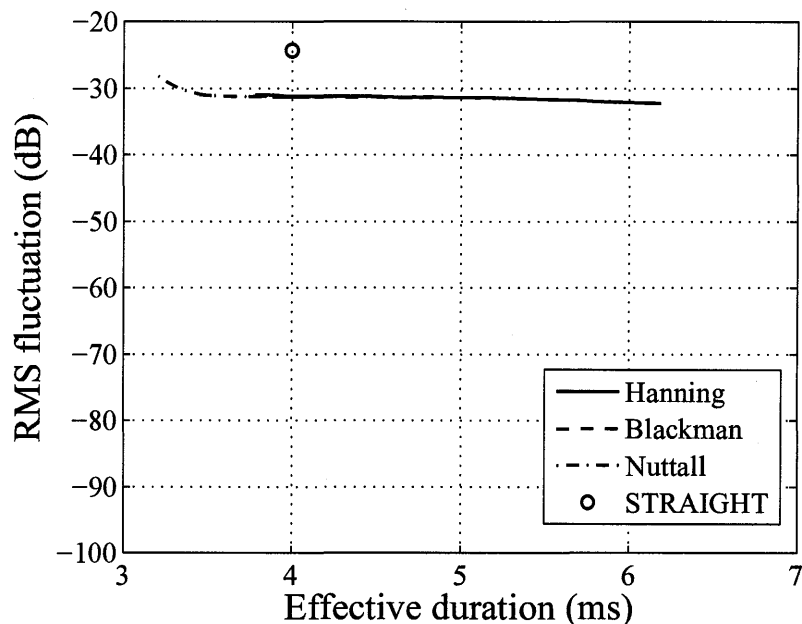


図 2.20 窓関数の種類・長さで時間変動との関係. SNR20 dB における結果.

合干渉項が相殺されず、窓関数の違いによる優劣無く持続時間と時間変動とが強い相関を持つこととなる。すなわち、単一の窓関数を用いた短時間フーリエ変換の場合と同様に調波間内に全てのメインローブが含まれる Hanning 窓では数%の誤差では影響されなかったことを意味する。

2.4.7 音声分析に適した窓関数についての考察

これまでの計算機シミュレーションにより、TANDEM 窓における窓関数長と時間変動との関連を明らかにし、TANDEM が単一の窓関数を用いた短時間フーリエ変換・STRAIGHT における相補的時間窓と比較した場合、時間分解能・時間変動両方に関して優れていることを示した。周期信号に雑音が存在する場合においても同様の評価を行い、相補的時間窓よりも耐雑音性に優れていることを示した。さらに、基本周波数の推定誤差が性能に与える影響についても評価を行った。雑音が存在しない場合では、Nuttall 窓と Blackman 窓の時間変動は、STRAIGHT の相補的時間窓と同様の持続時間において、相補的時間窓より小さい時間変動となる。そのため、Blackman 窓、Nuttall 窓が TANDEM 窓に適していると考えられる。雑音が存在する場合、窓関数長の増加に対して時間変動は SNR よ

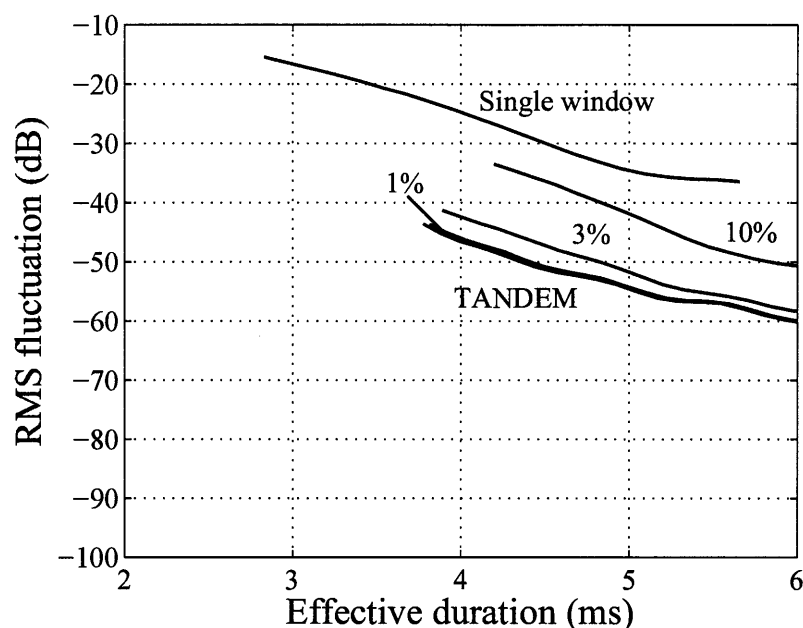


図 2.21 基本周波数の推定誤差が TANDEM 窓の持続時間に与える影響. Hanning 窓の場合

り 10 dB 程度低い数値までは減少するが、それ以上窓関数長を増加させても時間変動の改善は望めない。実環境で収録した音声の SNR は一般的に 60 dB を超えることは無いため、窓関数長を増加させた Blackman 窓、Nuttall 窓を用いたとしても、大きな時間変動の改善は見込めない。

また、音声分析では声帯振動が毎回微細に異なることより、高い時間分解能が要求される。声帯振動の特性は 1 振動毎に異なり、振動間隔も短時間で微細に変動する。さらに雑音も存在することから、時間変動の低減よりも時間分解能を優先したほうが品質の向上を見込める可能性がある。基本周波数の推定誤差は全ての窓関数において TANDEM の優位性を失い、窓関数の持続時間にのみ依存した時間変動を与えることとなる。これらのことから、相補的時間窓よりも高い時間分解能で -60 dB 程度の時間変動となる、調波の約 2.4 倍の長さを持つ Blackman 窓や、調波の約 3 倍の長さを持つ Nuttall 窓が TANDEM 窓には適していると考えられる。

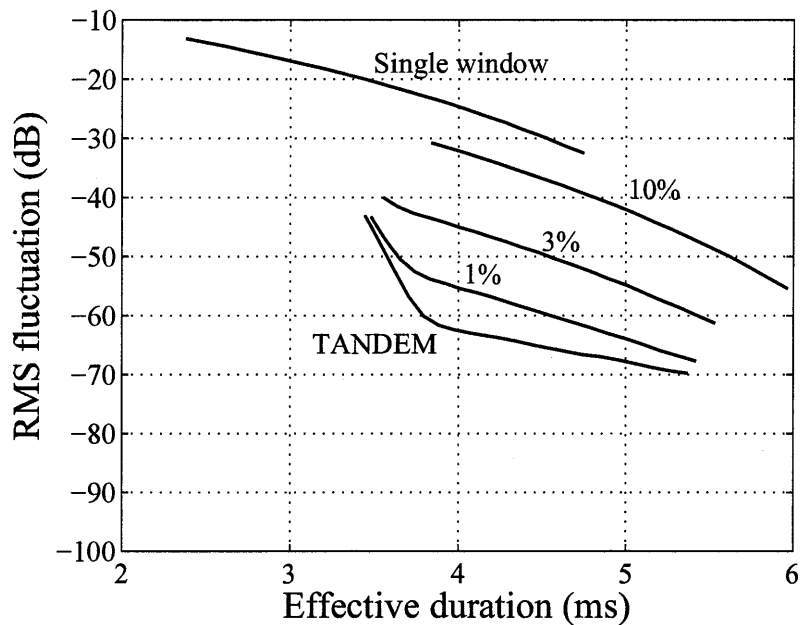


図 2.22 基本周波数の推定誤差が TANDEM 窓の持続時間に与える影響. Blackman 窓の場合

2.5. まとめ

本章では、分析時刻に依存しない周期信号の分析法 TANDEM を提案した. TANDEM は、任意のインパルス応答と任意の周期をもつパルス列とのたたみ込みで表現される周期信号より、基本周波数の整数倍において、インパルス応答本来のパワーを保存することができる. また、分析時刻がパワーに与える影響を、従来法である STRAIGHT 分析より低減できることを示した.

はじめに、任意の信号を時間的に繰り返し配置することで周期信号とみなした場合、周波数軸でどのような影響を及ぼすかについて説明を行った. 周期信号には周期性の影響が含まれることを例示し、その原因が、周期信号を有限の時間の窓関数で切り出しであることを示した. 窓関数を用いて周期信号を任意の時刻で切り出すと、窓関数と周期信号との相対位相が原因となり、パワースペクトルが分析時刻に依存することを明らかにした.

これらの問題点を解決するため、時間シフトした 2 つの窓のパワースペクトルを加算する簡便な方法 TANDEM の提案を行い、その原理について説明を行った. TANDEM では、サイドローブの小さい窓関数であれば任意の窓関数を選択できるため、条件を見た

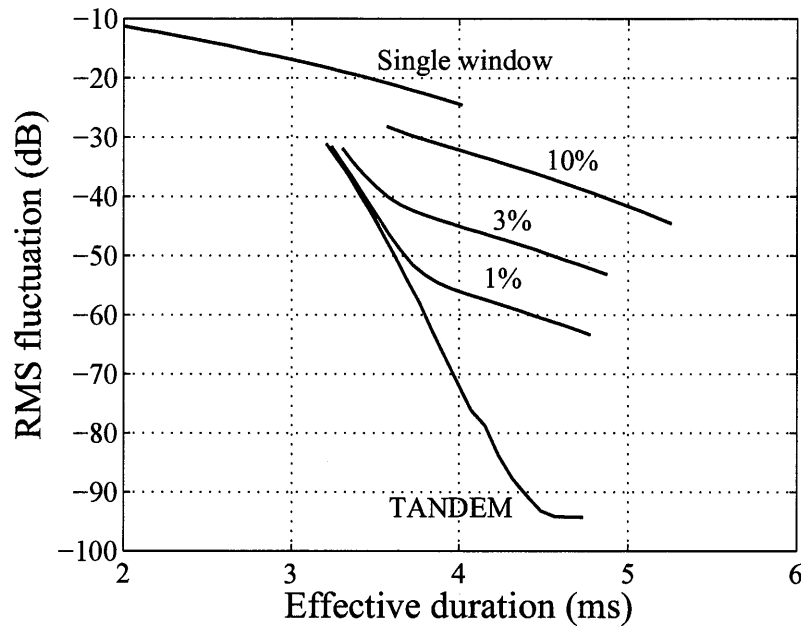


図 2.23 基本周波数の推定誤差が TANDEM 窓の持続時間に与える影響. Nuttall 窓の場合
す 3 種類の窓関数を候補として, 計算機シミュレーションで分析時刻への依存性を指標と
した評価を行った.

単一の窓関数を用いた短時間フーリエ変換, STRAIGHT における相補的時間窓を比較
対象として, 様々な条件の下で TANDEM が優れていることを示した. まず, 雑音も周期
性の乱れも存在しない周期信号で計算機シミュレーションを行い, TANDEM は, 音声分
析に要求される高い時間分解能を保ちつつ分析時刻による変動成分を従来法より抑圧で
きることを明らかにした. 雑音が存在する状況でも同様の計算機シミュレーションを行
い, TANDEM が耐雑音性の観点でも従来法より優れていることを示した. さらに, 推定
された周期信号の周期に誤差が含まれている場合についても評価を行い, 周期の推定誤
差に対しても頑健であることを示した.

これらの計算機シミュレーションの結果に基づき, TANDEM に適した窓関数の選定に
ついて考察し, その結果, 音声分析を行う場合の TANDEM に適した窓関数は Blackman
窓, Nuttall 窓という結論が得られた.

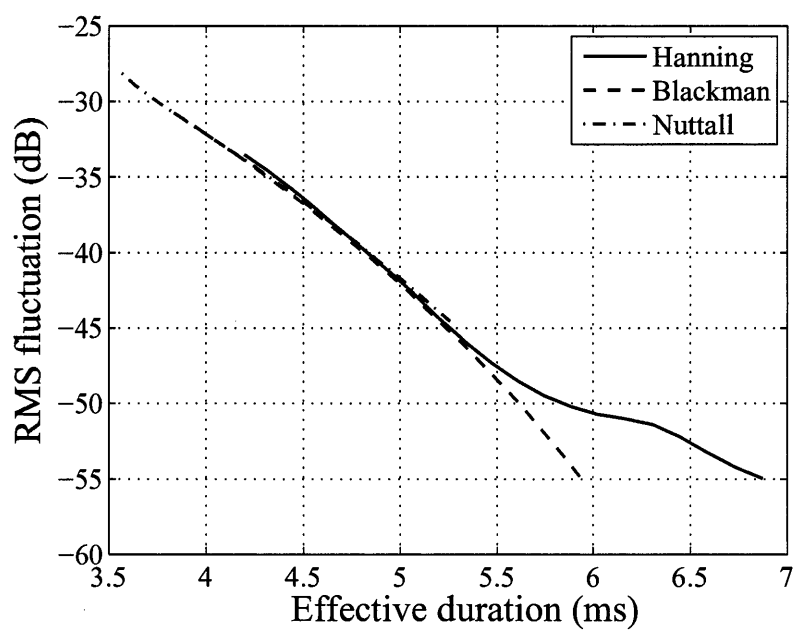


図 2.24 基本周波数に 10 %の誤差含まれる場合における TANDEM 窓の持続時間と時間変動.

第 3 章 音声を用いたクロススペクトル法によるインパルス応答推定法

3.1. はじめに

マイクロフォンで収録した音声には，口元から観測点までの伝達関数が含まれる．本論文では，口の形や舌の位置による調音フィルタ $V(\omega)$ と口元から観測点までの伝達関数 $H(\omega)$ ，有声音ならば，声帯振動のスペクトル情報 $G(\omega)$ も含めてスペクトル包絡 $P(\omega)$ としている．マイクロフォンの設置位置を人間の口元にするすることで，反射・残響といった伝達関数 $H(\omega)$ の影響は直接音に比べ十分小さくできる．しかし，口元から遠くなるほど伝達関数 $H(\omega)$ の影響が大きくなり，合成音の品質に影響することが考えられる．口元から観測点までの伝達関数を高精度に推定することが出来れば，スペクトル包絡 $P(\omega)$ をより正確に推定できる可能性がある．

人間の発話した音声がどのように放射されるかを測定する研究は，古くから行われている．放射音圧に関しては，Flanagan⁵³⁾ が等身大のマネキンの口内にスピーカを設置し測定を行っている．電話の品質評価のための疑似口やダミーヘッドでは，口からの放射と頭部および体による反射・回折が考慮されている^{60, 61)}．しかし，実際に人間の口から放射される音は，口唇だけではなく鼻孔や咽頭壁からも放射される^{62, 63)}．また，人間の体形や頭部は，個人毎に形状が異なる．声道内部における音圧分布は母音毎に異なっているため^{64, 65)}，実際に発話した音声の放射パターンは，個人性，音韻依存性を持つこととなる．先行研究では，クロススペクトル法を用いた音声発話時の頭部周辺におけるインパルス応答推定法を提案し，音韻依存性を確認している⁵⁶⁾．

しかし，人間の音声をクロススペクトル法の入力とした場合，低域と高域におけるパワーの不足による SNR の低下という問題があった．そのため，文献⁵⁶⁾ で提案した測定法では，低域の SNR に関しては解決法を提案しているが，10 kHz 以上の成分を低域通過フィルタにより除去していた．

本研究では，物理系において，ある時間に存在するエネルギーは指数関数的に減衰す

る特徴を持ち、インパルス応答は様々な周波数の減衰振動の和として表されること、そして、口唇から放射される音は体を回折して伝わることから、回折しやすい低域の応答は長時間存在し、回折しにくい高域の応答は短い時間で消滅するという特徴に着目した。一般的に、インパルス応答は、残響時間や暗騒音に埋もれる時間までを切り出し評価を行っているが、これらの特徴を考えると、切り出したインパルス応答の後半部分には直接音に起因する高い周波数成分は含まれておらず、暗騒音に起因する成分のみが存在することとなる。そのため、インパルス応答に起因する成分を周波数毎に取り出すことが出来れば、高域の不要成分を除去できる可能性がある。

本章は、はじめに音声を入力としたクロススペクトル法によるインパルス応答推定法について説明を行い、測定されたインパルス応答から基本周波数以下の補正と反射成分を抑圧する後処理について述べる。同一条件で複数回測定されたインパルス応答における特性の標準偏差を調査することで、この後処理を行った特性においても高域になるに従い信頼性が低下している(本論文ではばらつきが大きいことを信頼性が低いと表現する)ことを示す。次いで、高域における信頼性が低い原因を、収録音声の SNR、物理系におけるインパルス応答の特徴から明らかにする。直接音成分に起因する時間周波数領域を抽出することを狙い、時間軸の伸縮と低域通過フィルタを用いることによる高域の信頼性向上について検討する。時間軸の伸縮を離散系で実装する際生じる誤差の補償を狙い、振幅周波数特性、群遅延特性における誤差を明らかにする。この誤差を補償することで元信号に与えていた影響が改善出来ることを示す。同一条件下で複数回測定したインパルス応答の標準偏差を評価指標とし、提案手法を用いることで高域における信頼性が向上していることを明らかにする。

3.2. インパルス応答の推定手法

インパルス応答の推定法として、M 系列信号⁶⁷⁾ や TSP (Time-Stretched Pulses)⁶⁸⁾ を用いた手法がある。これらの手法は、単一パルスを用いた同期加算法⁷³⁾ より高い SNR でインパルス応答を測定することが可能であるが、測定用信号は予め用意する必要がある。そのため、人間の音声そのもののインパルス応答を推定する際にこれらの手法を利用することはできない。そこで、入力に対する制限が少ないクロススペクトル法^{73, 70, 72)} により音声のインパルス応答測定を行う。クロススペクトル法とは図 3.1 に示すように、入力波形と出力波形の伝達系を推定する手法である。この方法では、まず $x(n)$ 、 $y(n)$ の波形をある長さのブロックに分割し、各ブロックを任意の窓関数により切り出しを行う。

次に、各ブロックにフーリエ変換を施し、パワースペクトルおよびクロススペクトルを計算し、全てのブロックの平均処理を行うことで伝達関数を推定できる。具体的には以下の式によって推定を行う。

$$\hat{H}(k) = \frac{\overline{X^*(k)Y(k)}}{\overline{X^*(k)X(k)}} \quad (3.1)$$

ここで、 $H(k)$ は推定対象となる伝達関数、 $\hat{H}(k)$ は推定された伝達関数、 $Y(k)$ は出力信号の DFT、 $X(k)$ は入力信号 (推定用信号) の DFT、 X^* は X の共役複素数、 $\overline{}$ は X の多数回平均、 k は離散周波数番号を表す。

文献⁵⁶⁾では、図 3.2 のように人間の音声をクロススペクトル法の入力に使用する実例を示している。また、ダミーヘッドの口から放射されるインパルス応答を、M 系列信号を用いて推定した場合と、実際に人間が発声した音声を入力としてクロススペクトル法を用いた場合とについて比較しており、どちらの手法を用いても音声の主要な周波数成分程度までは問題なく使用できる精度が得られていることが示されている。以下では、文献⁵⁶⁾で用いたクロススペクトル法を用い音声を入力としたインパルス応答の推定法と、後処理について説明を行う。

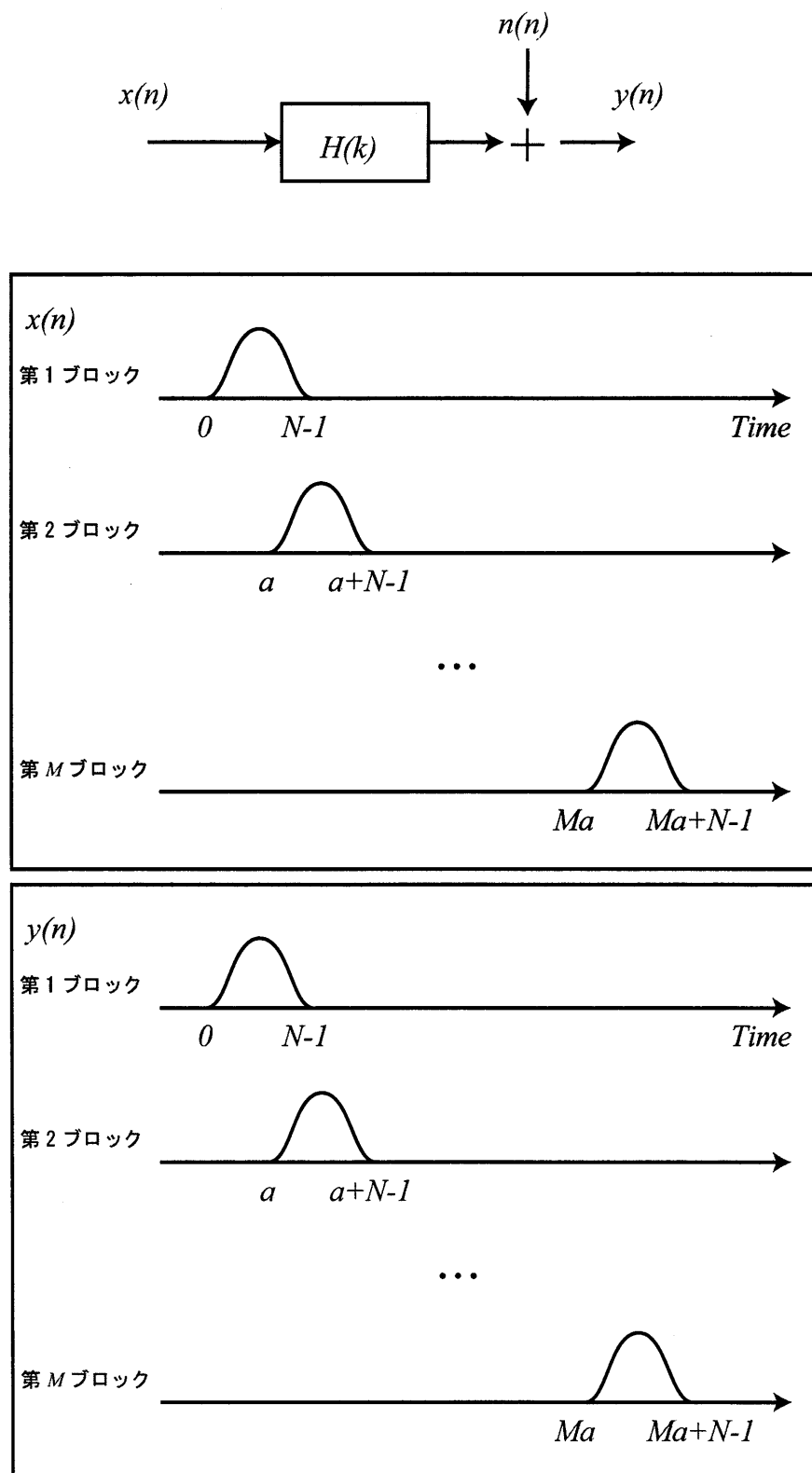


図 3.1 クロススペクトル法における窓掛け

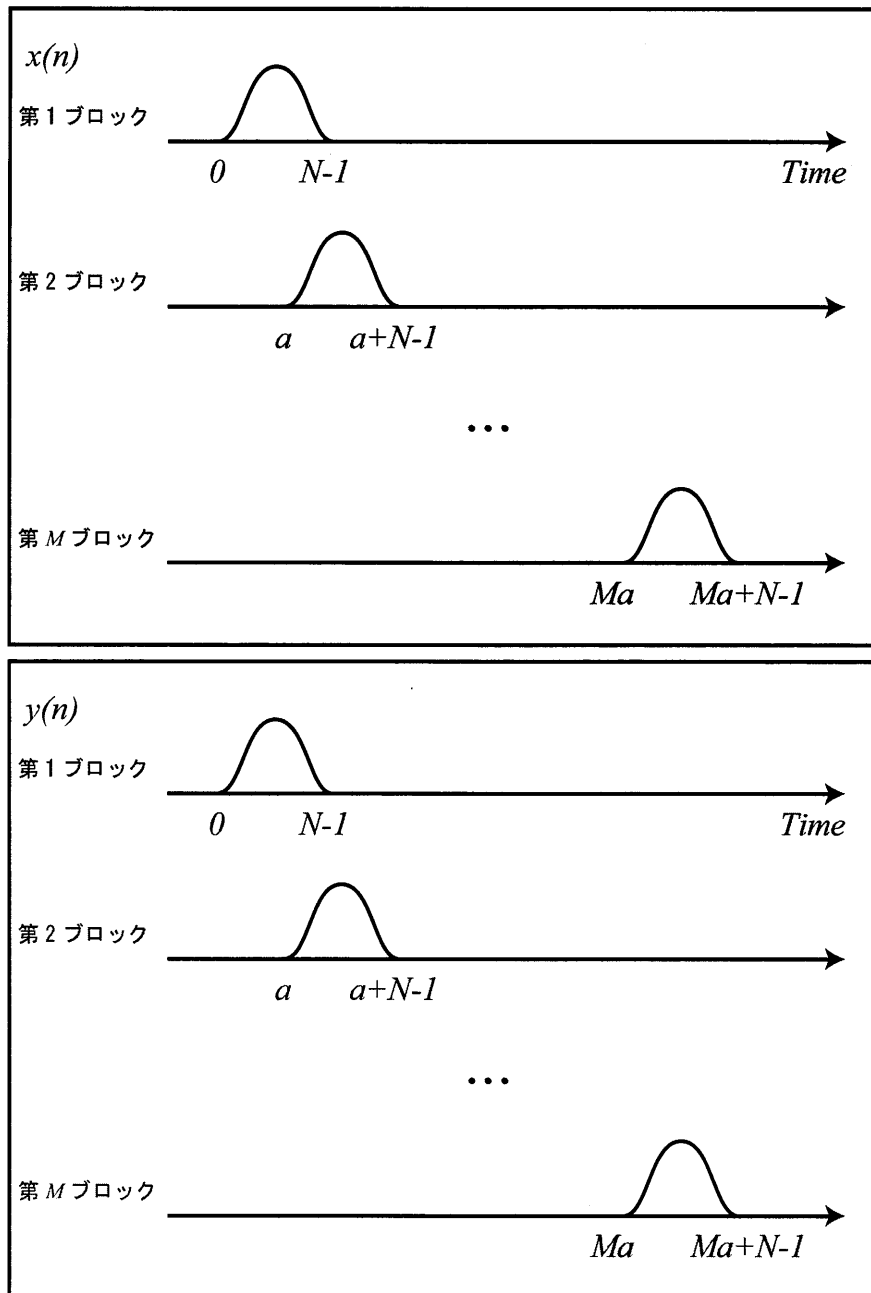
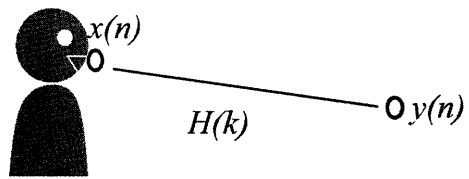


図 3.2 音声を入力とした場合のクロススペクトル法

3.2.1 音声の収録

口元と左外耳道入り口へマイクロフォンを設置し、クロススペクトル法の入力、出力信号を得るために被験者(男性1名)が発声している音声を収録した。収録は無響室で行い、サンプリング周波数は44.1 kHz、量子化ビット数は24bitとした。また、マイクロフォンは無指向性のコンデンサマイクロフォンであるDPAの4060BMを用いた。口元のマイクロフォンは呼気の吹き掛けを避けるため、口元よりやや左へ外れた位置に設置した(図3.3)。被験者には、無響室の中央で立った状態で発声を行うよう教示した。音声は、母音/a/のみとし、基本周波数を上下に変動させ発声してもらった。これは、クロススペクトル法における入力信号に対する制約条件「パワースペクトルに零点が存在しない」を満たすためである。1つの発声の持続時間は8秒程度とし、発声の間には数秒の休止を挿入した。収録時間は休憩を挟まずに30分程度とし、音声は合計186回収録した。

図3.4に収録音声と暗騒音の1/3オクターブバンド毎のエネルギーを示す。音声は1発声分(6秒程度)、暗騒音は発声毎に挟む休止区間(4秒程度)を分析したものである。収録音声の平均基本周波数は約170 Hzであった。図3.4より音声の存在する帯域においては、高いSNRが得られていることが分かる。一方、100 Hz以下と10kHz以上においてSNRが低下していることも分かる。

3.2.2 インパルス応答の測定手順

ここでは、クロススペクトル法により求められたインパルス応答から、基本周波数以下の補正と、反射成分を除去する処理の手順を説明する。クロススペクトル法に用いる窓関数は8192点(約186 ms)のBlackman窓とした。クロススペクトル法で用いる窓関数による誤差は文献^{70, 72)}により検討されており、Blackman窓は誤差が大きいと指摘されている。しかし、本研究では、入力信号としてダイナミックレンジの大きな音声を利用しているため、サイドローブによる影響が少ないBlackman窓を選択した。音声を測定用信号に用いたクロススペクトル法によるインパルス応答推定に適した窓関数については、次章で詳しく検討する。

基本周波数以下の補正

クロススペクトル法を用いて測定されたインパルス応答を図3.5に示す。図3.4からも明らかなように低域と高域はSNRが低いため暗騒音による影響を受けており、インパル

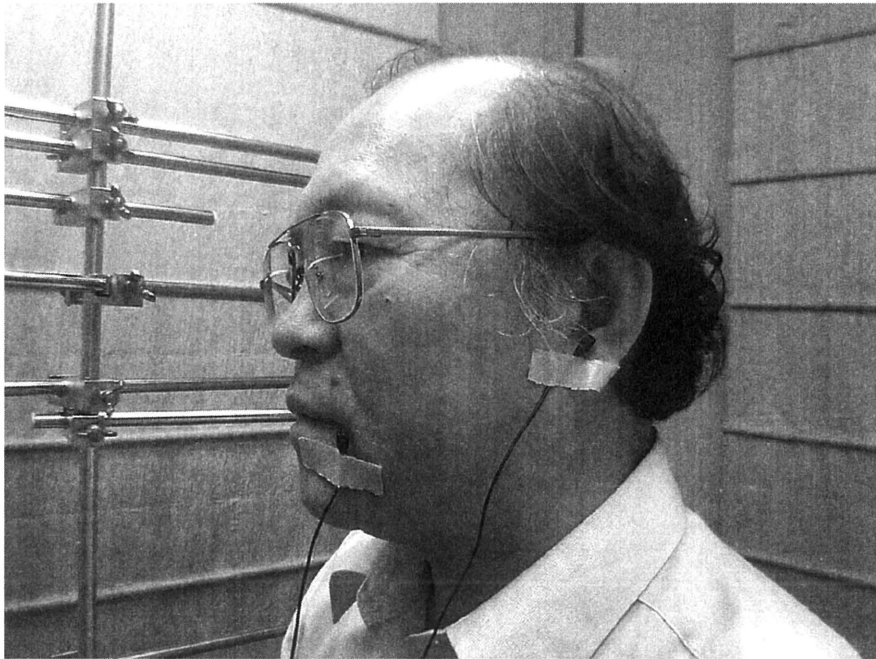


図 3.3 マイクの設置位置について

ス応答では、大きな低域の雑音成分が正負双方向に広がっている。

ここで、0 Hzにおける位相差が0になるという拘束条件を利用し、基本周波数以下の特性を、基本周波数における値から求めている。具体的には、周波数0における位相を0とし、絶対値は基本周波数における値と一致させ、間を線形補間によって求めた。この処理は、高いSNRの値を用いて低いSNRにおける情報を補うものである。この処理を行ったインパルス応答を図3.6に示す。低域における非因果的な応答は大きく抑圧されている。

反射波の抑圧

補正処理により、低域における非因果的な成分を抑圧することができた。しかし、インパルス応答には、反射波の影響が含まれている。ここで、最も早く到来する反射波の時刻以降の応答が含まれないよう、シグモイド関数によって抑圧する。また、主応答より手前の応答に関しても、暗騒音に埋もれる時刻より手前をシグモイド関数により抑圧する。今回の収録において最も早く到来する反射波は膝からの反射であり、耳元までの距離は約70 cmであった。そのため、主応答の到来時間から反射波が到来する4.1 ms(約140

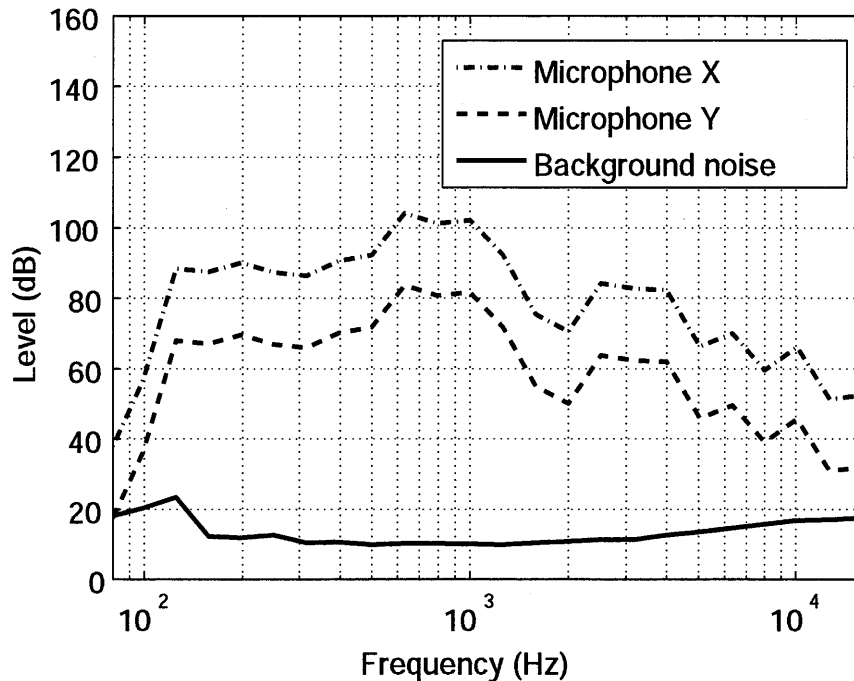


図 3.4 収録音声と暗騒音の 1/3 オクターブバンド毎のエネルギー

cm) 以降の応答を抑圧した。

クロススペクトル法により得られたインパルス応答と、低域の補正、反射波の影響を取り除いたインパルス応答の振幅周波数特性を図 3.7 に示す。(a) が処理前の振幅周波数特性、(b) が処理後の振幅周波数特性である。処理前のインパルス応答と比較すると、反射成分に起因する細かい変動成分が取り除かれ、滑らかな形状が得られていることが分かる。

3.2.3 多数回測定することによる特性のばらつき

インパルス応答の直接音成分は、同一条件において複数回測定しても不変である。ここでは、複数回測定したインパルス応答の振幅周波数特性と位相特性における標準偏差を求めることで、帯域毎の信頼性を明らかにする。ただし、位相特性に関しては、位相特性をそのまま評価に用いるのではなく、位相特性の周波数の導関数である群遅延にパワースペクトルを重み付けした関数を用いた。なお、パワースペクトルで重み付けした群遅延は、平均遅延時間となる¹⁴⁾。

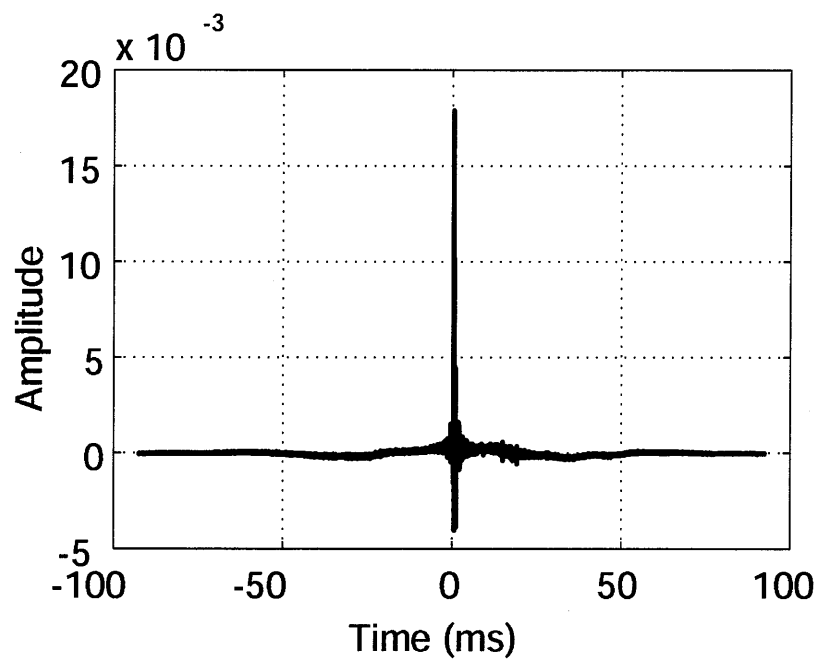


図 3.5 クロススペクトル法により測定されたインパルス応答

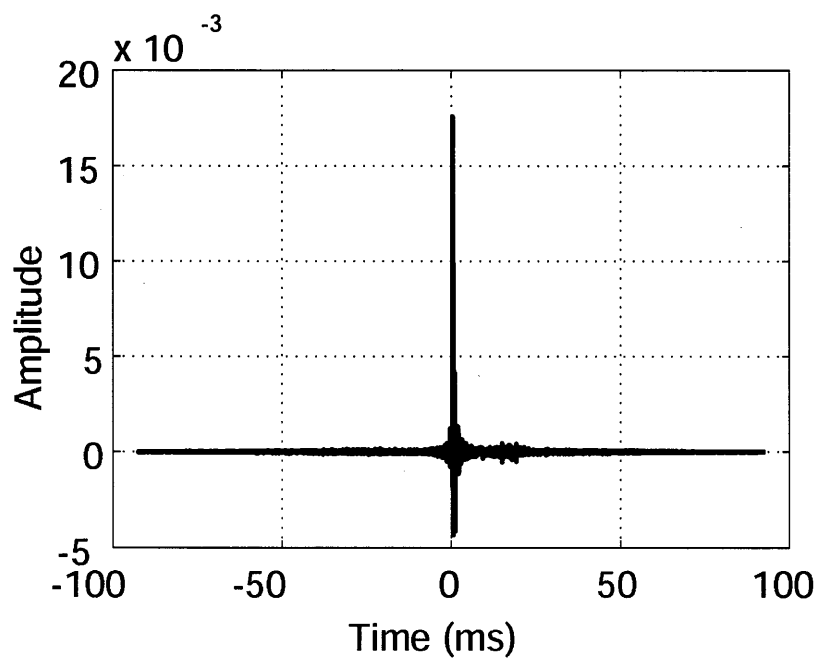


図 3.6 低域の補正処理を行ったインパルス応答

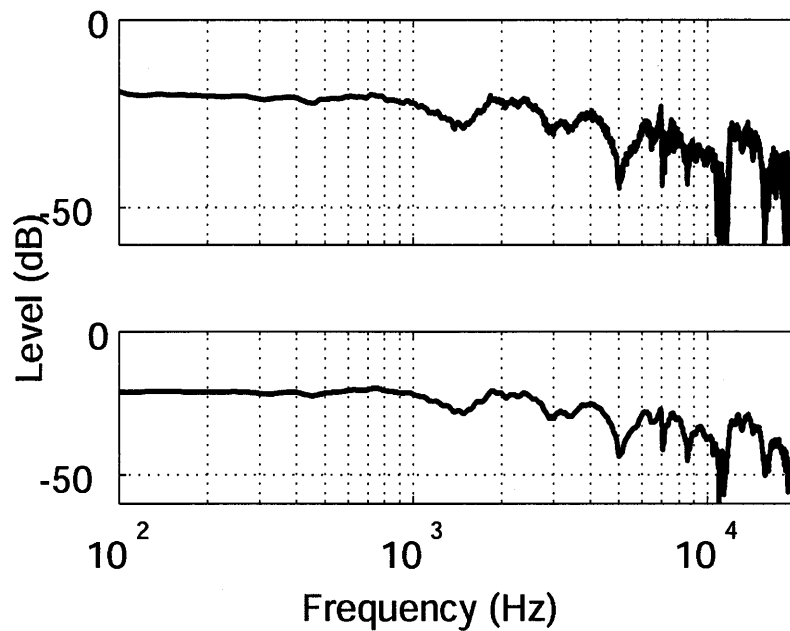


図 3.7 測定されたインパルス応答の振幅周波数特性 (a) と，低域の補正処理，反射波の抑制を行ったインパルス応答の振幅周波数特性 (b)．

インパルス応答は，長時間マイクを設置することによるマイク位置の移動による特性の変動を避けるため，基本周波数から 1 kHz までのパワーが等しくなるよう正規化した．図 3.8 に，正規化したインパルス応答から求めた振幅周波数特性の平均と標準偏差，パワースペクトルで重み付けした群遅延特性の平均と標準偏差を示す．振幅周波数特性，重み付き群遅延特性共に，高域になるに従い標準偏差が増加する傾向にある．これは，高域における SNR の低下が原因であると考えられる．また，雑音の存在による影響は，2 乗平均となる振幅周波数特性よりも平均となる群遅延特性の方が大きい．以下の節では，直接音成分の時間周波数領域における特徴に着目し，高域の信頼性を向上させる手法の提案を行う．

物理系においてある時刻に存在するエネルギーは，一般的に指数関数的に減衰する特徴を持つ⁷²⁾．そのため，測定されたインパルス応答の直接音成分は，様々な周波数の減衰振動の和で表されることが出来る．また，耳元に到来する音は，直接到来する成分以外にも顔や体など全身からの回折が含まれる．これらの回折成分は，人間が音を発声する際必ず含まれる成分であり，壁や床からの反射のように抑圧してしまわない必要がある．回折は周波数によって異なる様相を呈する．例えば低域においては，回折し

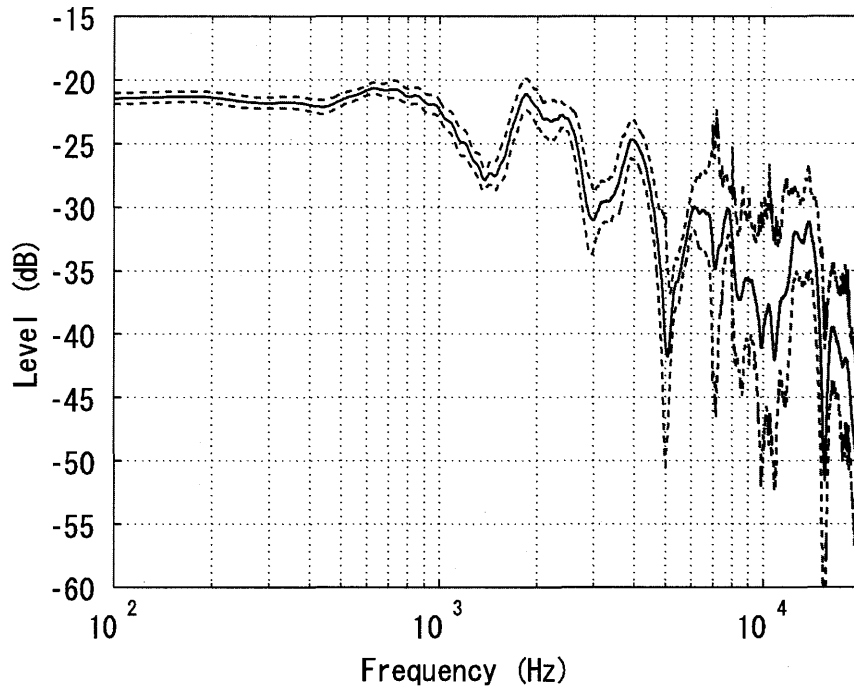


図3.8 エネルギーを正規化したインパルス応答 (処理後) の振幅周波数特性 (a) と、パワースペクトルで重み付けした群遅延特性 (b) の、平均 (実践) と標準偏差 (破線)

やすいため、応答は長時間持続すると考えられる。逆に高域においては、回折は起きにくいため、応答の持続時間は短いと考えられる。そのため、直接音に起因する成分における帯域毎の持続時間は、スピーカ振動系におけるインパルス応答同様に、周波数に反比例している可能性が考えられる。

この問題に対処する方法として、時間軸を対数関数で伸縮し、低域通過フィルタを通すことで、直接音に起因する時間周波数領域を選択する手法が提案されている⁷⁵⁾。時間周波数領域を選択するパラメタとして、周波数と持続時間が反比例であると仮定した場合の減衰率 ζ と、減衰を開始する時刻 t_s を用いる。文献⁷⁵⁾で提案している時間周波数領域を選択するための伸縮関数を以下に示す。まず、減衰開始時刻 ($0 < t < t_1 + t_s$) の範囲では、伸縮率が一定であり、上限の周波数が $f_s/2$ であることがわかっているの以下のようなになる。

$$\tau(t) = \frac{f_s t}{2f_N} \quad (3.2)$$

ここで f_s はサンプリング周波数、 f_N は時間軸伸縮後通過させる低域通過フィルタの遮断

周波数である．次に， $t \geq t_1 + t_s$ では，対数関数が次のようになる．

$$\tau(t) = \frac{\log(K)}{2\pi\zeta f_N} (\log(t - t_s) - \log(t_1)) + \frac{f_s(t_1 + t_s)}{2f_N} \quad (3.3)$$

ただし， $t_1 = \log(K)/(\pi f_s \zeta)$ であり， K は振幅が $1/K$ となる時刻である．今回は， ζ に対応する物理的実体はないが， $\zeta/\log(K)$ を周波数と，考慮すべき時間長との対応をつけるためのパラメタと解釈すれば良い．図 3.9, 3.10 に，線形時間軸上のインパルス応答，非線形伸縮された時間軸上のインパルス応答を示す．非線形伸縮された時間軸上では，原点に近い時刻が相対的に引き伸ばされている．そのため，非線形に伸縮された時間軸によって表現された信号を低域通過フィルタによって処理することは，低域に影響を与えることなく，原点から離れた高域の雑音を取り除く効果がある．離散時間における時間軸の伸縮は，伸縮後最も短いサンプリング間隔で全体をサンプリングし，補間することによって近似的に行う．ここで生じる誤差については，次節にて議論する．

最後に低域通過フィルタにより平滑化された信号を，逆関数を用いて元の時間軸に戻す．逆関数は， $0 < \tau < f_s(t_1 + t_s)/f_N$ の範囲では，

$$t(\tau) = \frac{2f_N\tau}{f_s} \quad (3.4)$$

となり，また， $\tau \geq f_s(t_1 + t_s)/f_N$ では，

$$\begin{aligned} t(\tau) &= t_1 \exp\left(\frac{\tau - C_2}{C_1}\right) + t_s \\ C_1 &= \frac{\log(K)}{2\pi\zeta f_N}, \quad C_2 = \frac{f_s(t_1 + t_s)}{2f_N} \end{aligned} \quad (3.5)$$

となる．

3.2.4 離散系で実装することによる誤差と，誤差の補償

減衰開始時刻 t_s までの応答は，時間軸伸縮処理において全ての周波数が低域通過フィルタの遮断周波数 f_N Hz 以下の周波数になるように時間軸を引き伸ばされるため，誤差は生じない．ただし，時間軸の伸縮を離散系で実装する場合，時間軸伸縮後最も狭い間隔で全体をサンプリングしなおし，間を線形補間で求めることによる誤差が生じていると考えられる．仮にパルスの時間軸を線形に引き伸ばし，間を線形補間した場合，引き伸ばした時間軸上では三角波となる．ここで三角波が方形波同士の畳み込みであることを考えると，

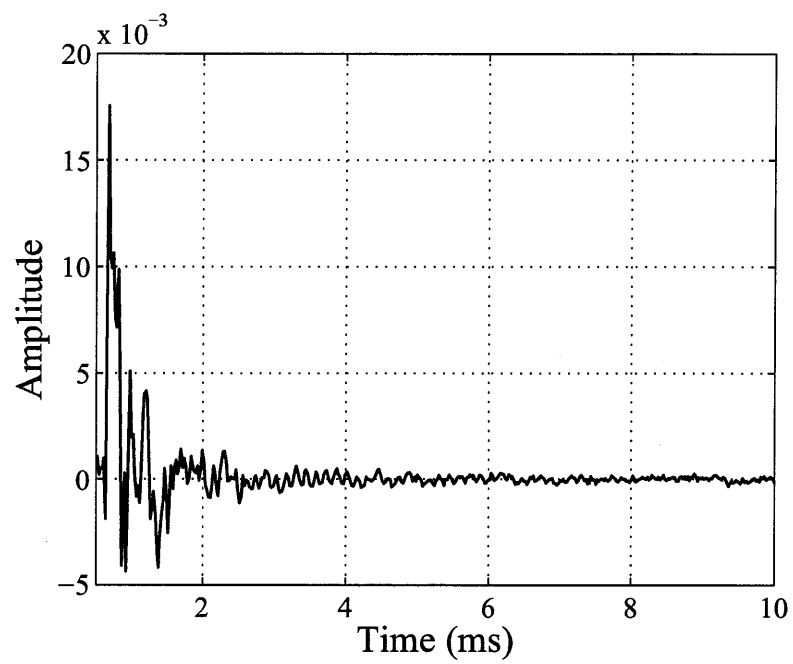


図 3.9 線形の時間軸で観測したインパルス応答

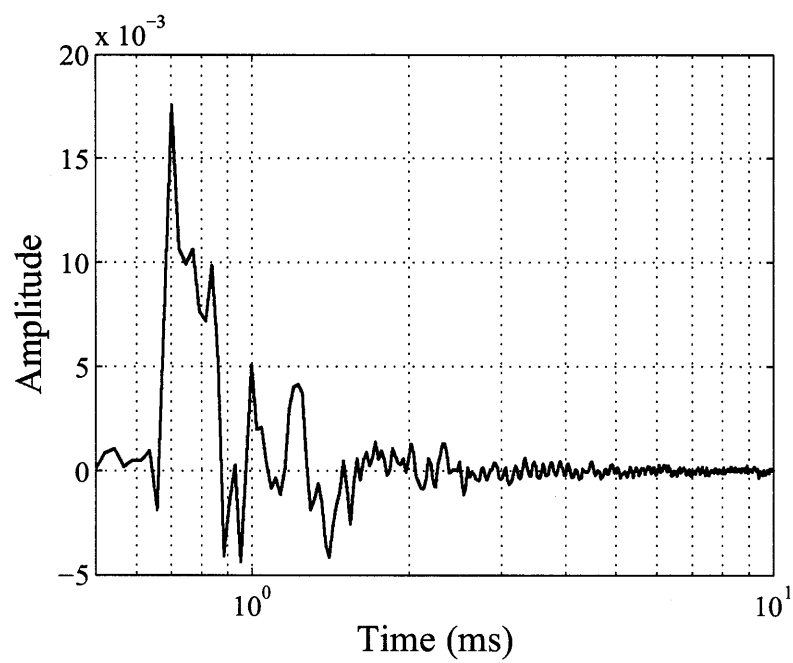


図 3.10 非線形伸縮された時間軸で観測したインパルス応答

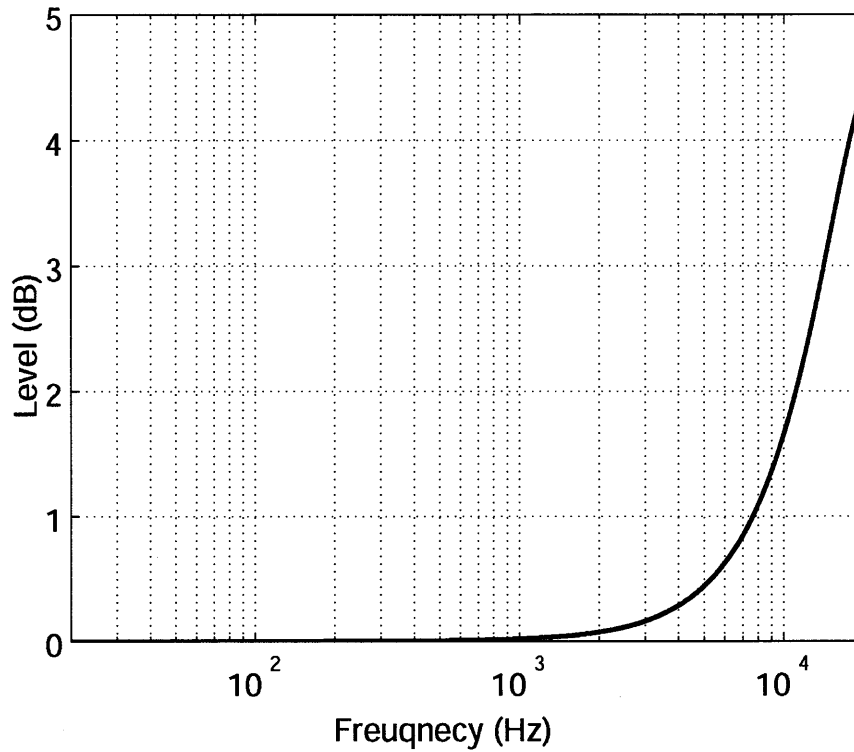


図 3.11 時間軸伸縮処理前後の応答から求めた振幅周波数特性の誤差

三角波の振幅周波数特性は方形波の振幅周波数特性 $\sin(\omega T_s)/\omega T_s$ の自乗となり、高域になるに従いパワーが減衰する。この三角波に低域通過フィルタを通した後に時間軸を復元するため、時間軸伸縮処理後の信号は、高域のパワーが不足している。ここでは、これらの要因によって生じる振幅周波数特性および群遅延特性の誤差について説明する。入力信号を $x(t)$ 、出力信号を $y(t)$ とした場合、振幅周波数特性の誤差は、 $E(k) = |X(k)|/|Y(k)|$ となる。また、群遅延特性の誤差は、 $\psi'_e(k) = \psi'_x(k) - \psi'_y(k)$ である。

入力信号 $x(t)$ をパルスとし求めた誤差を、振幅周波数特性について図 3.11 に示す。振幅周波数特性の誤差は高域になるに従い増加しており、10kHz 以上で 3dB を超える誤差が生じていることが分かる。一方群遅延は、有効数字 12 桁において全帯域の誤差が 0 であったため、誤差は含まれないと判断した。

この誤差を補償するためには、処理後のインパルス応答 $y(t)$ の振幅周波数特性に、パルスを入力としたときの誤差を付与すればよい。つまり、補償後の振幅周波数特性は $|Y(k)||E(k)|$ となる。処理後のインパルス応答と、処理後に誤差の補償を行ったインパルス応答とを、RMS 誤差により比較する。評価に用いる信号は、原点から減衰開始時刻

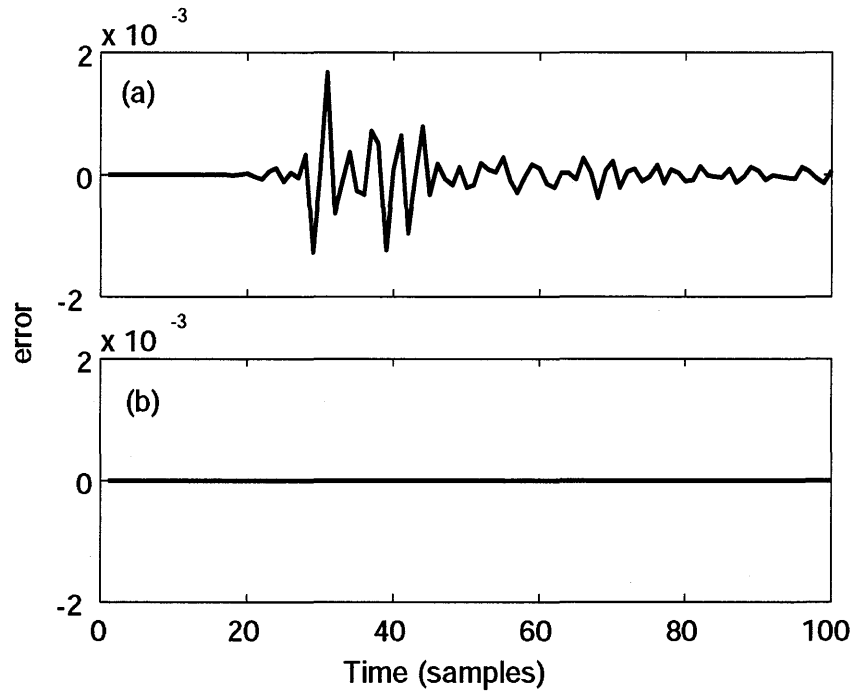


図 3.12 補償前 (a) と補償後 (b) の誤差信号.

t_s までの応答とした．まず，誤差補償前と誤差補償後の誤差信号を図 3.12 に示す．上図は誤差補償前，下図は誤差補償後である．誤差は，補償後のほうが明らかに少ないことが分かる．低域の補正と反射波の抑圧を行ったインパルス応答を時間軸伸縮処理し，RMS 誤差を算出したところ 2.45×10^{-4} であった．また，誤差保証後の RMS 誤差は 5.79×10^{-10} であった．補償前に比べ大幅に改善されており，問題なく使用できる精度が得られていることが分かる．

3.3. 評価

提案した時間軸伸縮と低域通過フィルタを組み合わせた手法による信頼性の向上について評価を行う．ここでは，時間軸伸縮のパラメタである減衰率 ζ を，初期反射が到来する時刻において基本周波数成分まで抑圧されるように設定した．評価指標は，振幅周波数特性 (dB) における標準偏差と，パワースペクトルで重み付けした群遅延特性の標準偏差，持続時間の標準偏差である．図 3.13 に振幅周波数特性の平均と標準偏差を，図 3.14 にパワースペクトルで重み付けした群遅延特性の平均と標準偏差を示す．どちらの図も

上段は低域補正と反射波の抑圧処理を行った特性，下段はさらに提案手法を施した特性である．10 kHz 以上の標準偏差の平均は，振幅周波数特性で処理前の 80%，重みつき群遅延特性では処理前の 33.8% まで低減していた．

パワースペクトルで重み付けした群遅延特性と持続時間については，高域における信頼性が向上していることが分かる．これは，提案手法によって主応答の到来時間から離れた暗騒音が抑圧されているためと考えることができる．一方，振幅周波数特性に関しては，他ほど大きな改善は見られない．これは，群遅延，持続時間はエネルギー中心から離れた位置に存在する雑音に大きく影響されるのに対し，パワースペクトルは相対的に影響されにくいからであると考えられる．

3.4. まとめ

本報告では，クロススペクトル法で測定された人間の発声するインパルス応答において，高域における信頼性を向上させる手法を提案し，評価を行った．まず，音声をクロススペクトル法の入力とした場合低域と高域の SNR が低くなり，信頼性が低くなることを実測により明らかにした．口から放射される音の回折を考慮した帯域毎の持続時間が周波数に反比例する可能性を述べ，直接音に起因する時間周波数領域を選択することで，高域における信頼性が改善できることを示した．そこで，時間軸の伸縮と低域通過フィルタを組み合わせ，直接音に起因した時間周波数領域を取り出す手法について説明した．また，離散系で実装する際に用いる補間関数によって誤差が生じることを示し，その誤差を補償する手法を明らかにした．同一条件の 186 回の発声データを用いて提案手法とこれまでの手法の評価を行った．その結果，群遅延特性，持続時間に関しては信頼性が大幅に向上することが示された．また，振幅周波数特性に関しては改善はあるものの，群遅延特性，持続時間ほどではないことが示された．

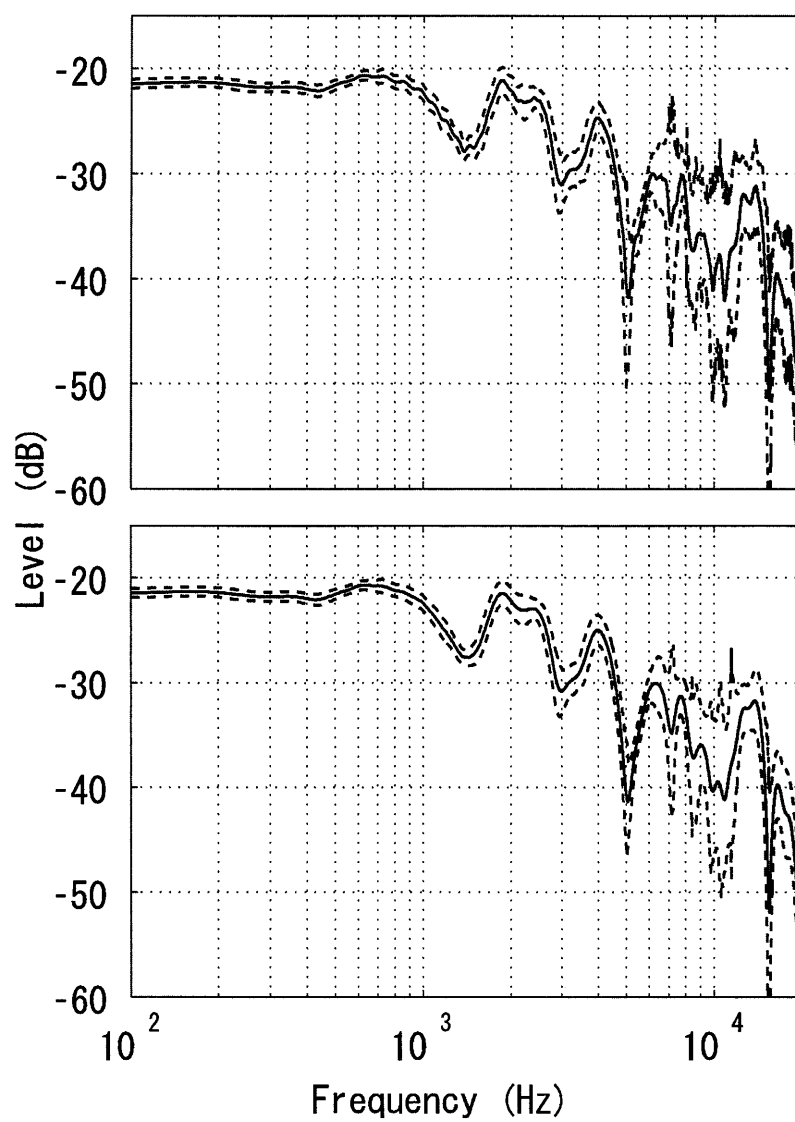


図 3.13 振幅周波数特性の平均 (実線) と標準偏差 (破線). 上段は処理前. 下段が処理後.

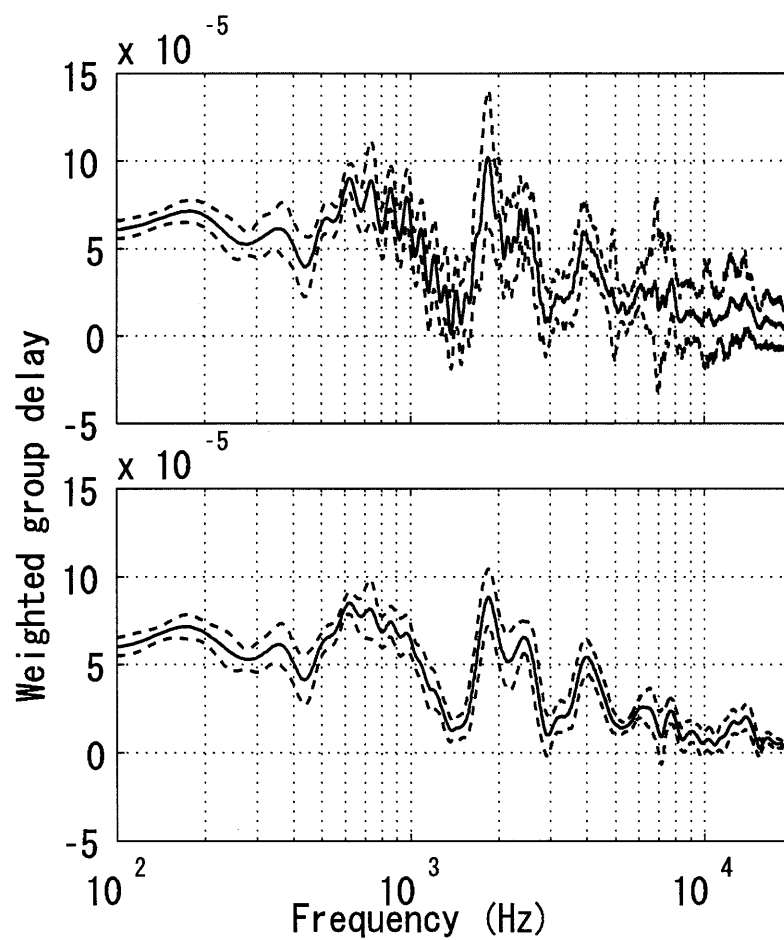


図3.14 パワースペクトルで重み付けした群遅延特性の平均(実践)と標準偏差(破線). 上段は処理前, 下段が処理後.

第 4 章 音声を用いたクロススペクトル法に適した時間窓の検討

4.1. はじめに

前章で示した音声の伝達関数推定はクロススペクトル法に基づいた方法である。クロススペクトル法では測定用信号を有限の時間窓により切り出すため、時間窓の種類と時間窓長に依存した推定誤差が含まれる^{69, 70)}。この推定誤差は、推定されたインパルス応答波形の振幅が変形することで生じる⁷¹⁾。クロススペクトル法を用いたインパルス応答推定では、時間窓長をインパルス応答長 (本論文では指数関数的減衰のインパルス応答が 60dB 減衰するまでの時間をインパルス応答長とする) の 4 倍より長く設定し、方形波窓以外を用いれば良いとされる⁷⁰⁾。推定されたインパルス応答の振幅変形が小さい時間窓も提案されている⁷²⁾。しかし、従来の研究で行われてきた検討は、測定用信号としてホワイトノイズを用いた議論である。音声を測定用信号とした場合における推定誤差と、これらの要因との関係は示されていない。

音声を測定用信号とした場合では、音韻により測定用信号の特性が異なるため、文献⁷²⁾のように、入力信号に対し最適な時間窓を設計することは困難であると考えられる。そこで本章の目的は、クロススペクトル法の測定用信号を音声とした場合、推定誤差に影響を与える要因を示し、推定誤差の小さい時間窓の条件を明らかにすることである。はじめに、ホワイトノイズ、音声、音声と同様の振幅周波数特性を持つ雑音、ピンクノイズを用いてシミュレーションを行うことで、測定用信号の周期性、振幅周波数特性の傾斜、狭い帯域のダイナミックレンジが推定誤差に与える影響を調査する。なお、本論文におけるダイナミックレンジとは、音声のホルマントに起因するスペクトル包絡のピーク間におけるエネルギーの最大値と最小値との差とし、振幅周波数特性の傾斜に起因する全帯域でのエネルギー差とは区別する。これらの検討に基づき、測定用信号のダイナミックレンジが推定誤差に影響することを示す。任意のダイナミックレンジを持つ雑音を用いてシミュレーションを行い、ダイナミックレンジと推定誤差との観点から推定誤差

の小さい時間窓の条件を明らかにする。

測定用信号のダイナミックレンジが 30 dB 以下ならば、リース窓の推定誤差が最も小さいことが示された。音声のように、狭いピーク間隔でダイナミックレンジが 40 dB 以上となるような測定用信号ならば、サイドローブの小さい時間窓が適していることも明らかとなった。

4.2. クロススペクトル法を用いたインパルス応答推定と時間窓による振幅変形

この節では、クロススペクトル法に用いる時間窓により生じる振幅変形⁷¹⁾について説明する。クロススペクトル法は、以下の式のように系への入力信号とそれに対する出力信号とのクロススペクトルを計算することで、インパルス応答を推定する手法である。

$$\hat{H}(k) = \frac{\overline{X^*(k)Y(k)}}{\overline{X^*(k)X(k)}} \quad (4.1)$$

ここで、 $H(k)$ は推定対象となる伝達関数、 $\hat{H}(k)$ は推定された伝達関数、 $Y(k)$ は出力信号の DFT、 $X(k)$ は入力信号 (推定用信号) の DFT、 X^* は X の共役複素数、 \overline{X} は X の多数回平均、 k は離散周波数番号を表す。クロススペクトル法では対象となる信号の時間波形を有限長の時間窓により切り出す。このとき、切り出された信号の時間波形は、用いた時間窓に依存して振幅が変形する。そのため、推定されたインパルス応答波形は、振幅変形による誤差が含まれる。文献⁷¹⁾では、この振幅変形による誤差を検討し、以下の式のように定義している。

$$d(n) \equiv \frac{\sum_{m=0}^{N-1-n} w(m)w(m+n)}{\sum_{m=0}^{N-1} w(m)^2} \quad (4.2)$$

$$\hat{h}(n) = d(n)h(n) \quad (4.3)$$

$w(n)$ は切り出しに用いる時間窓、 $h(n)$ は推定対象となるインパルス応答、 $\hat{h}(n)$ は推定されたインパルス応答、 $d(n)$ は時間窓による振幅変形係数である。式 4.3 は、推定されたインパルス応答の振幅が $d(n)$ により歪むことを示している。また、時間窓長 N をインパルス応答長 M より長くすることで、 $w(m)w(m+n)$ がほぼ $w(m)^2$ とみなせる範囲にインパルス応答主要部が含まれることにより振幅変形の影響が低減できることを示している。

しかし、音声のような信号を用いる場合、式 4.3 以外にも測定用信号のダイナミックレンジや周期性、振幅周波数特性の傾斜に起因する誤差が生じる可能性もある。本論文では、測定用信号として、ホワイトノイズ、音声、音声と同様の振幅周波数特性を持つ雑音、ピンクノイズ、ダイナミックレンジを操作した雑音を用いて、測定用信号と推定誤差との関係を調べる。それらの結果に基づいて、測定用信号を音声とした場合のインパルス応答推定に適した時間窓の選択を行う。

4.3. 計算機シミュレーションによる評価

4.3.1 推定対象とするインパルス応答

本研究の目的は、クロススペクトル法を用いて口から測定点までのインパルス応答を推定することである。そこで、推定対象となるインパルス応答（以下では真のインパルス応答と呼ぶことにする）を、前章の収録により求めた口元から外耳道入り口までの母音/a/におけるインパルス応答とした。推定対象となるインパルス応答の時間波形を図 4.1、振幅周波数特性を図 4.2 に示す。インパルス応答推定には、声帯の振動周波数を上下に連続的に変動させた約 8 秒の音声を 186 個用いた。1 発話分の振幅周波数特性の例を図 4.3 に示す。なお、収録に用いたサンプリング周波数は 44100 Hz である。

4.3.2 ホワイトノイズ、音声との対比

はじめに、ホワイトノイズと無響室で収録した音声とを測定用信号として用い、インパルス応答推定のシミュレーションを行う。音声は、無響室で収録した基本周波数を変動させた持続母音/a/である。推定されたインパルス応答は、正規化誤差エネルギー E_r を用いて評価する。正規化誤差エネルギー E_r は、推定されたインパルス応答と真のインパルス応答との差を表す誤差の指標である。正規化誤差エネルギー E_r は、次式で表される。

$$E_r = \frac{\sum_{n=0}^{M-1} \{h(n) - \hat{h}(n)\}^2}{\sum_{n=0}^{M-1} h(n)^2} \quad (4.4)$$

ここで、 $h(n)$ は真のインパルス応答、 $\hat{h}(n)$ は推定されたインパルス応答である。シミュレーションに用いた時間窓は過去の文献⁷⁰⁾と同様の方形波窓、Hanning 窓、Hamming

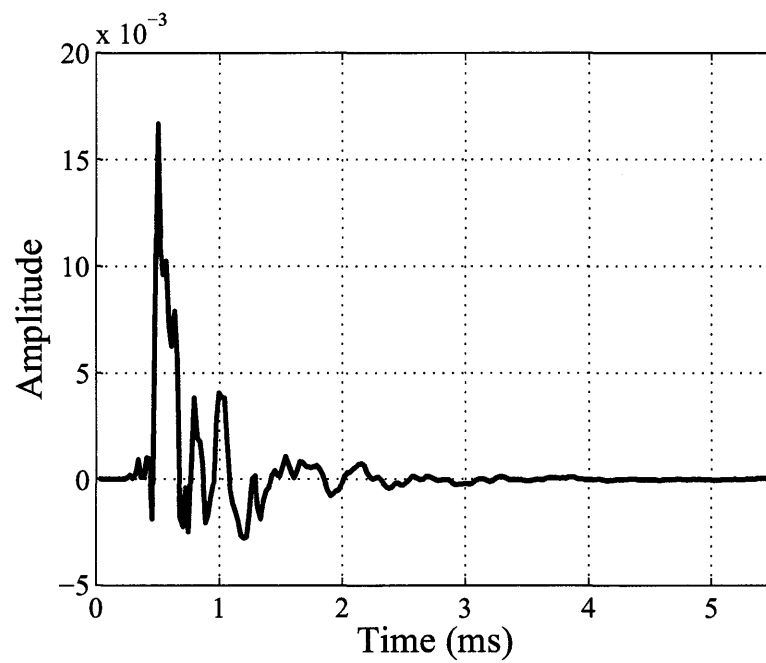


図 4.1 推定対象となるインパルス応答の時間波形.

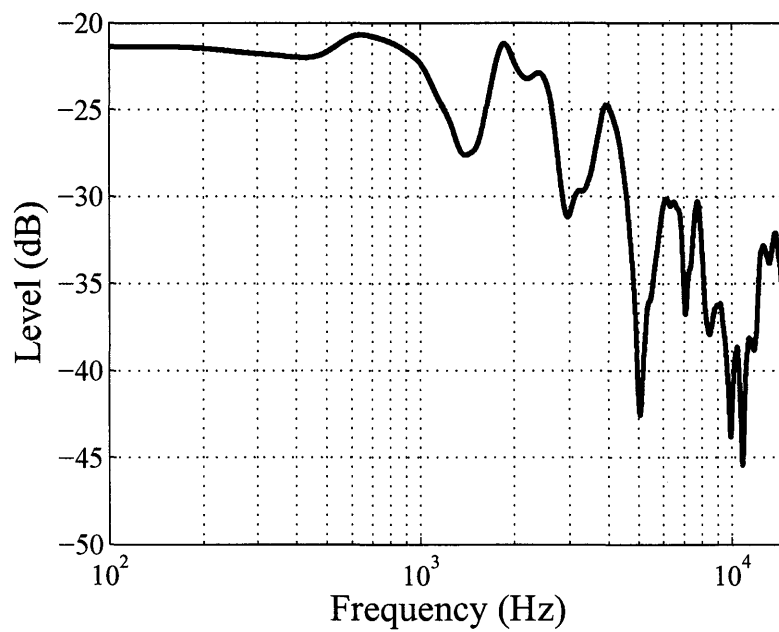


図 4.2 推定対象となるインパルス応答の振幅周波数特性.

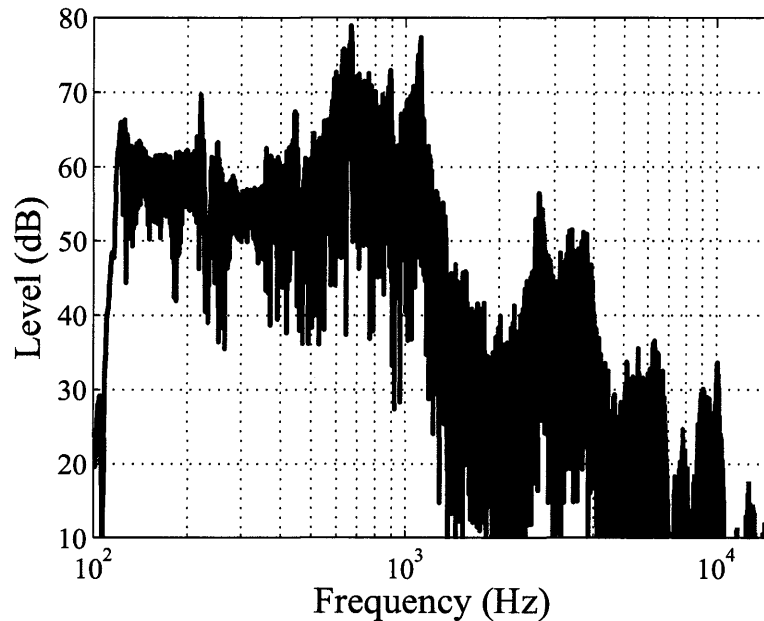


図 4.3 インパルス応答推定に用いた音声の振幅周波数特性.

窓, Blackman 窓, Riesz 窓, そしてホワイトノイズにおいて最適な時間窓とされる $\sin^{0.62}$ 窓⁷²⁾ の 6 種類である. 時間窓長はインパルス応答長の 1, 2, 4, 8, 16 倍に設定し, スペクトル計算のための平均回数は 10^4 回とした. 本論文で行う全てのシミュレーションは, 測定用信号以外の条件を等しくした.

測定用信号としてホワイトノイズを用いた結果を図 4.4 に, 測定用信号として音声を用いた結果を図 4.5 に示す. 図 4.4 より, 方形波窓と Blackman 窓の推定誤差が他の時間窓より大きくなることが分かる. 一方, 音声を測定用信号とした場合, N/M が 8 以上において Hanning 窓, Blackman 窓の推定誤差が他の時間窓より小さくなる. 推定誤差に差が生じた要因は, 音声の特徴であるダイナミックレンジ, 周期性, 振幅周波数特性の傾斜である. そこで, 音声と同様の振幅周波数特性を持つ雑音と, ピンクノイズとを測定用信号としてシミュレーションを行う. この音声有色雑音は, 図 4.3 に例として示した振幅周波数特性を全ての音声について平均し, ホワイトノイズに周波数領域で乗じたものを用いた. 音声による結果と音声と同様の振幅周波数特性を持つ雑音との比較をすることで, 音声の持つ周期性が推定誤差に与える影響について示す. ピンクノイズの結果と比較することで, 音声のダイナミックレンジ, 振幅周波数特性の傾斜が推定誤差に与える影響を明らかにする.

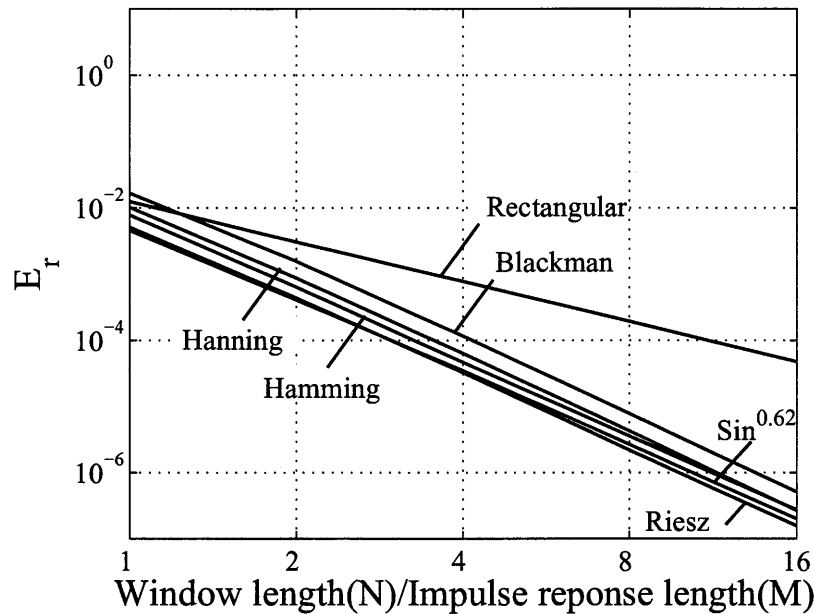


図 4.4 正規化誤差エネルギーと時間窓長 N とインパルス応答長 M の比との関係。測定用信号はホワイトノイズ。

4.3.3 音声と同様の振幅周波数特性を持つ雑音、ピンクノイズによるシミュレーション

測定用信号を音声とし、インパルス応答長の 8 倍以上の長さの時間窓を用いた場合、Hanning 窓、Blackman 窓による推定誤差が他の時間窓より小さくなることが示された。この節では、測定用信号として図 4.3 に例として示した音声と同様の振幅周波数特性を持つ雑音と、ピンクノイズとを用いて同様のシミュレーションを行う。同じ振幅周波数特性を持つ音声と雑音との結果を対比することにより、推定誤差に与える影響が音声の周期性か振幅周波数特性かを示す。音声と同様の振幅周波数特性を持つ雑音とピンクノイズとの結果を対比することにより、推定誤差に影響を与える要因が振幅周波数特性の狭い帯域のダイナミックレンジであるか、振幅周波数特性の傾斜であることを明らかにする。

図 4.6 に測定用信号を音声と同様の振幅周波数特性を持つ雑音とした場合における結果、図 4.7 に測定用信号をピンクノイズとした場合における結果を示す。測定用信号が音声と同様の振幅周波数特性を持つ雑音の結果と図 4.5 に示した音声の結果とを対比すると、推定誤差に違いがあるものの、 $N/M = 16$ において推定誤差の小さい時間窓は、Hanning

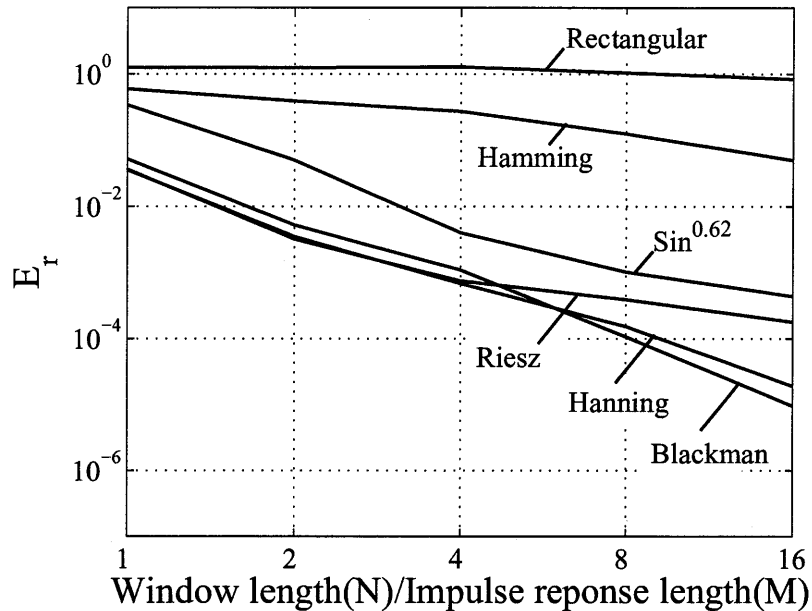


図 4.5 正規化誤差エネルギーと時間窓長 N とインパルス応答長 M の比との関係. 測定用信号は音声.

窓, Blackman 窓である. すなわち, 音声の周期性が推定誤差に与える影響は, 振幅周波数特性に比べ小さいと考えられる. 一方, 測定用信号をピンクノイズとした場合の結果では, ホワイトノイズと同様に, 方形波窓と Blackman 窓以外の推定誤差が小さい.

これらの結果より, 推定誤差に影響を与える要因は, 測定用信号のダイナミックレンジであると考えられる. ただし, ピンクノイズのように全帯域におけるダイナミックレンジが大きくとも, 狭い帯域においてダイナミックレンジが小さいような雑音では, ホワイトノイズと同様の傾向となる. すなわち, 狭い帯域でダイナミックレンジが大きくなるような測定用信号において, Hanning 窓, Blackman 窓の推定誤差が小さくなる. そこで, 振幅周波数特性のピーク間隔とダイナミックレンジを任意に設定できる雑音を用いてシミュレーションを行う. ダイナミックレンジと推定誤差との関係を各時間窓について明らかにする.

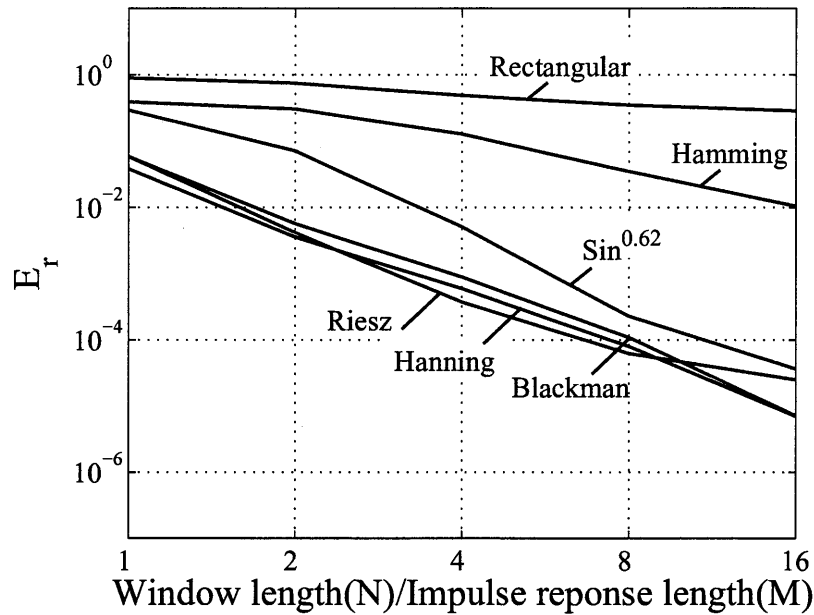


図 4.6 正規化誤差エネルギーと時間窓長 N とインパルス応答長 M の比との関係。測定用信号は音声と同じ振幅周波数特性を持つ雑音。

4.3.4 ピーク間隔とダイナミックレンジを任意に設定した雑音によるシミュレーション

これまでのシミュレーションにより、測定用信号が狭い帯域で大きなダイナミックレンジを持ち、推定対象となるインパルス応答長より十分長い時間窓を用いることが可能な場合、Hanning 窓、Blackman 窓の推定誤差が小さいことを示した。ここでは、振幅周波数特性のピーク間隔とダイナミックレンジを任意に設定できる雑音を用いてシミュレーションを行い、ダイナミックレンジと推定誤差との関係を明らかにする。

測定用信号は、全極型の伝達関数を用いて整形した雑音を用いた。振幅周波数特性は断面積が一定の音響管⁷⁴⁾を仮定し、以下の式により計算した。

$$D(\omega) = \frac{1}{1 + (1 - 10^{(-\alpha/20)}) \cos\left(\frac{\omega}{\beta}\right)} \quad (4.5)$$

ω は角周波数、 α はダイナミックレンジ、 β はピーク間隔を表す。ここでは、ピーク間隔を 1000 Hz に固定した。これは、声道長を成人男性の一般的な値である 17.5 cm⁷⁴⁾ とした場合におけるピーク間隔である。ダイナミックレンジは 10 dB から 60 dB まで 10 dB

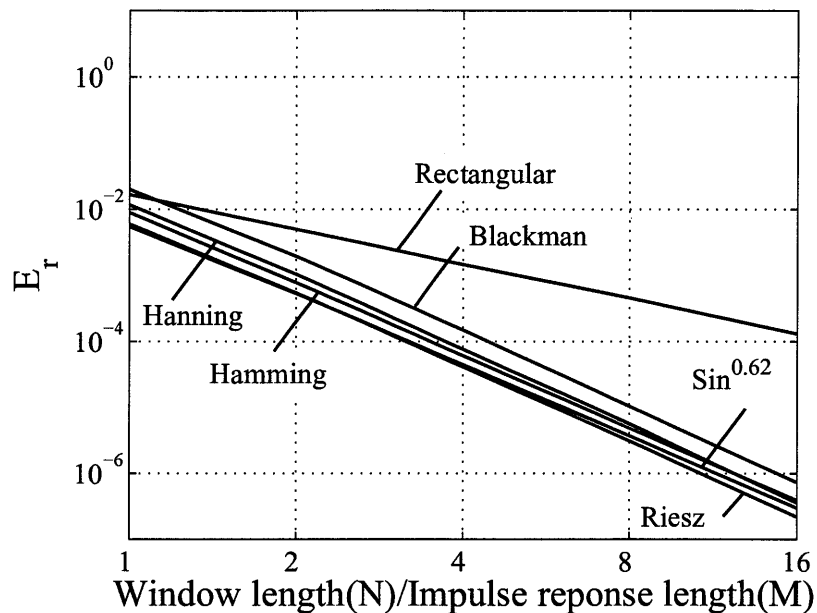


図 4.7 正規化誤差エネルギーと時間窓長 N とインパルス応答長 M の比との関係. 測定用信号はピンクノイズ.

毎に設定した. 図 4.8 はダイナミックレンジを 60 dB, ピーク間隔を 1000 Hz とした場合に得られた雑音の振幅周波数特性である.

結果を図 4.9, 4.10, 4.11, 4.12, 4.13, 4.14 に示す. まず, ダイナミックレンジが 10 dB の場合では, 全ての時間窓においてホワイトノイズにおける結果と同様の傾向が認められる. 方形波窓では, ダイナミックレンジが 30 dB 以上になると, 時間窓長が増加しても推定誤差がほとんど減少しない. Hamming 窓, $\sin^{0.62}$ 窓においても, ダイナミックレンジが大きくなるに従い, 方形波窓と同様の傾向となる. Riesz 窓は, ダイナミックレンジが 30 dB までは最も推定誤差の小さい時間窓である. しかし, ダイナミックレンジが 40 dB を超えると Hanning 窓よりも推定誤差が大きくなる. Hanning 窓, Blackman 窓は, ダイナミックレンジが小さい場合, Riesz 窓や $\sin^{0.62}$ 窓よりも推定誤差が大きい. ダイナミックレンジが 50 dB 以上になると, 他の時間窓より推定誤差が小さくなる.

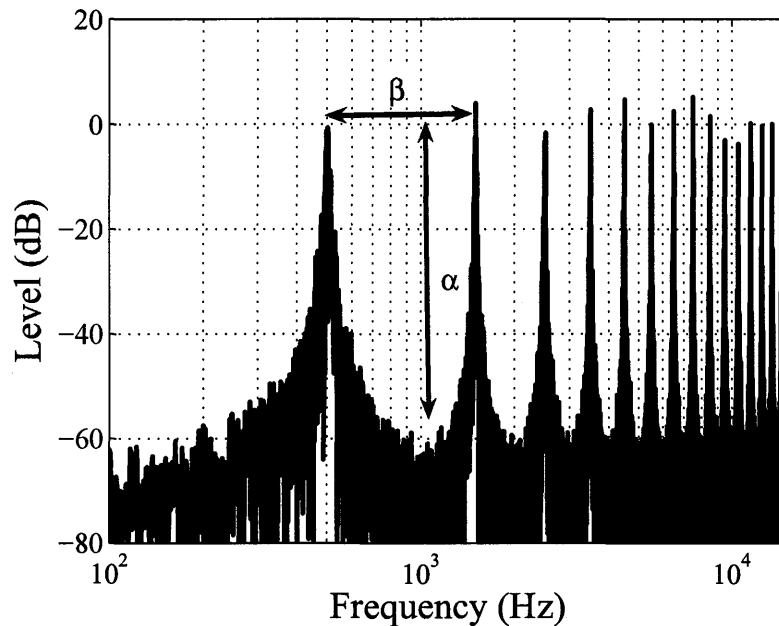


図 4.8 全極型の伝達関数を用いて整形した測定用信号の振幅周波数特性.

4.3.5 ダイナミックレンジに起因する推定誤差

ホワイトノイズのように平坦な振幅周波数特性を持つ信号を測定用信号とした場合、推定誤差は、式 4.2 に示す時間窓に依存した振幅変形である。しかし、測定用信号が平坦な振幅周波数特性を持たない場合、振幅変形以外にもダイナミックレンジに起因する推定誤差が含まれる。

ダイナミックレンジに起因する推定誤差は、時間窓のサイドローブが原因であると考えられる。振幅周波数特性が狭い帯域で大きく変化するような信号を、サイドローブの大きな時間窓により切り出した場合、振幅周波数特性のピークのエネルギーが、近傍の周波数に含まれる。そのため、ダイナミックレンジが大きい信号をサイドローブの大きな時間窓により切り出すと、振幅周波数特性の谷には、サイドローブに起因するエネルギーが多く含まれる。式 4.1 の分子と分母に含まれるサイドローブに起因するエネルギーは、必ずしも等しくはならない。そのため、分子と分母に含まれるサイドローブに起因するエネルギーの差が、推定誤差として含まれると推察できる。

図 4.15 に、 $N/M = 16$ におけるダイナミックレンジと推定誤差との関係を示す。推定誤差は、全ての時間窓において、ダイナミックレンジが増加するにしたがい増加する。サイド

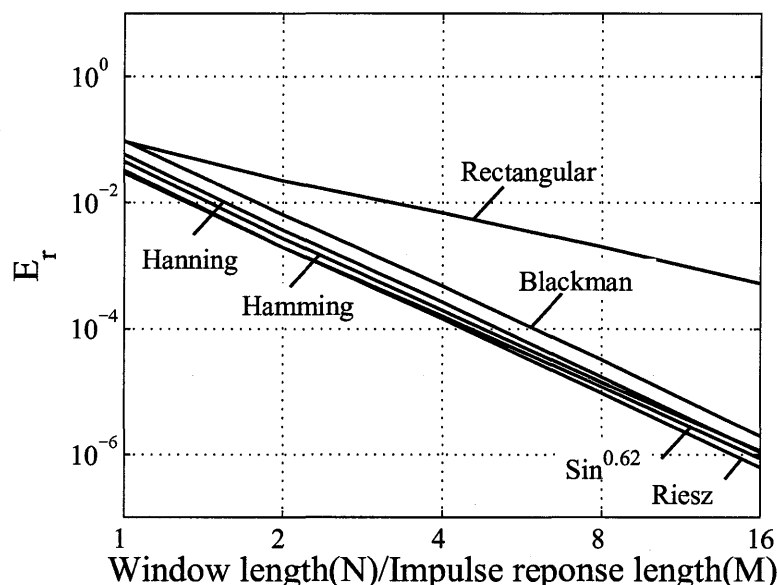


図 4.9 正規化誤差エネルギーと時間窓長 N とインパルス応答長 M の比との関係。測定用信号はダイナミックレンジ 10 dB の雑音。

ローブの大きな方形波窓，Hamming 窓， $\sin^{0.62}$ 窓は，サイドローブの小さい Hanning 窓，Blackman 窓と比較すると，小さいダイナミックレンジでも推定誤差が大きい．Hanning 窓と Blackman 窓とを比較すると，ダイナミックレンジが 60 dB 以下における推定誤差は Hanning 窓のほうが小さい．これは，振幅変形に起因する推定誤差のほうが，サイドローブに起因する推定誤差より大きいことを示唆している．

これらの結果は，真のインパルス応答の振幅周波数特性のダイナミックレンジによっても異なる可能性がある．真のインパルス応答の振幅周波数特性のダイナミックレンジが大きい場合，式 4.1 における $Y(k)$ にサイドローブの影響が含まれる．そのため，求める対象となるインパルス応答の振幅周波数特性のダイナミックレンジが大きい場合においても，Hanning 窓，Blackman 窓のようにサイドローブの小さな時間窓の推定誤差が小さくなると考えられる．これらについては今後の重要な検討課題である．

4.3.6 音声を測定用信号としたインパルス応答推定に適した時間窓

測定用信号のダイナミックレンジが 30 dB 以下ならば，Riesz 窓を用いることで推定誤差の小さいインパルス応答を得ることができる．ピーク間隔が 1000 Hz でダイナミック

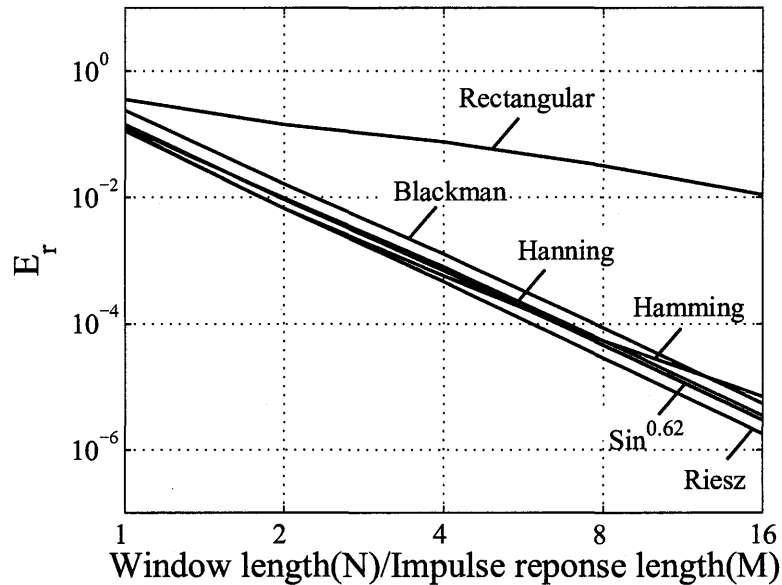


図 4.10 正規化誤差エネルギーと時間窓長 N とインパルス応答長 M の比との関係。測定用信号はダイナミックレンジ 20 dB の雑音。

レンジが 40 dB 以上の信号を測定用信号とした場合は、Hanning 窓、Blackman 窓を用いることで、他の時間窓より推定誤差を小さくできる。ダイナミックレンジが 60 dB を上回る場合では、Blackman 窓の推定誤差が最も小さく 40 dB から 60 dB にかけては Hanning 窓の推定誤差が最も小さい。一般的に音声は、フォルマント周波数により生じるピークの影響で、狭い帯域でダイナミックレンジが 40 dB 以上となる。そのため、音声を測定用信号としてインパルス応答推定を行う場合、Hanning 窓、Blackman 窓を用いれば良いといえる。

4.4. まとめ

本論文は、音声発話時の頭部周辺におけるインパルス応答推定を目的とし、クロススペクトル法の測定用信号を音声とした場合における、適切な時間窓を示した。ここでは、正規化誤差エネルギーを指標とし、ホワイトノイズ、音声、音声と同じ振幅周波数特性を持つ雑音、ピンクノイズ、任意のダイナミックレンジを持つ雑音について評価した。音声を測定用信号としたインパルス応答推定における推定誤差は、音声の振幅周波数特性における狭い帯域のダイナミックレンジに影響を受けることが示された。ダイナミックレ

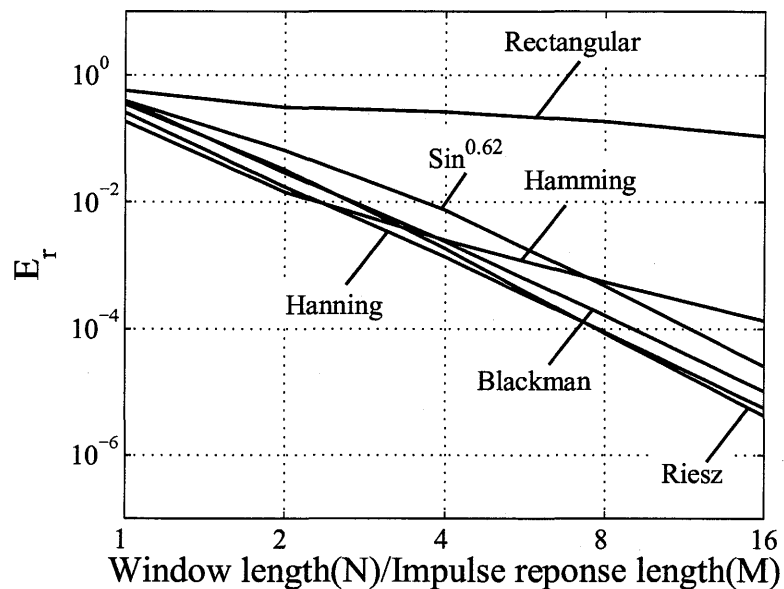


図 4.11 正規化誤差エネルギーと時間窓長 N とインパルス応答長 M の比との関係. 測定用信号はダイナミックレンジ 30 dB の雑音.

レンジが 30 dB 以下ならばリース窓を, 測定用信号のピーク間隔が 1000 Hz でダイナミックレンジが 40 dB 以上ならば Hanning 窓を, 60 dB 以上ならば Blackman 窓を用いれば良いことが明らかとなった. 狭い帯域で大きなダイナミックレンジを持つ音声を測定用信号としたインパルス応答推定に適した時間窓の条件は, サイドローブの小さい窓という結論が得られた.

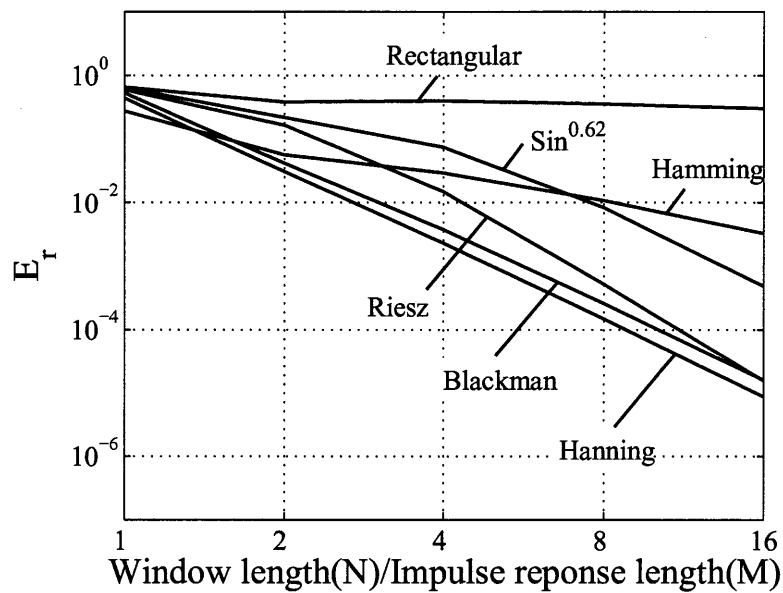


図 4.12 正規化誤差エネルギーと時間窓長 N とインパルス応答長 M の比との関係。測定用信号はダイナミックレンジ 40 dB の雑音。

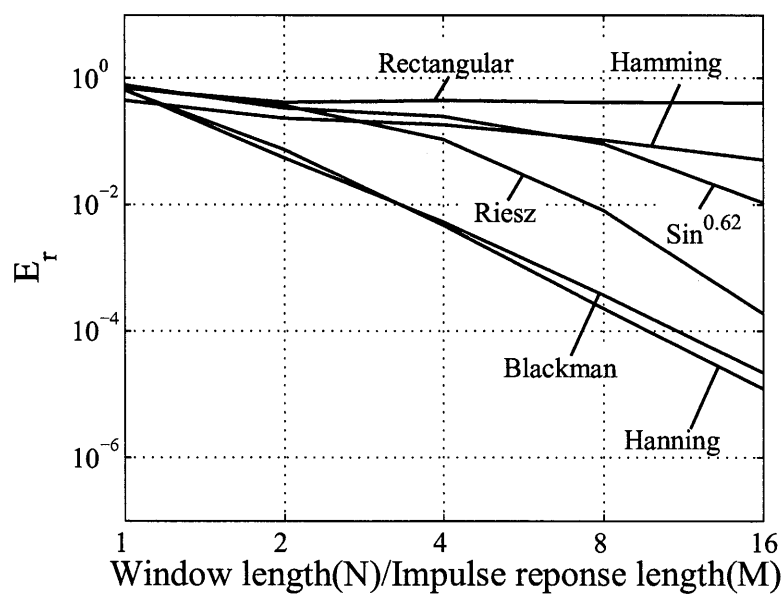


図 4.13 正規化誤差エネルギーと時間窓長 N とインパルス応答長 M の比との関係。測定用信号はダイナミックレンジ 50 dB の雑音。

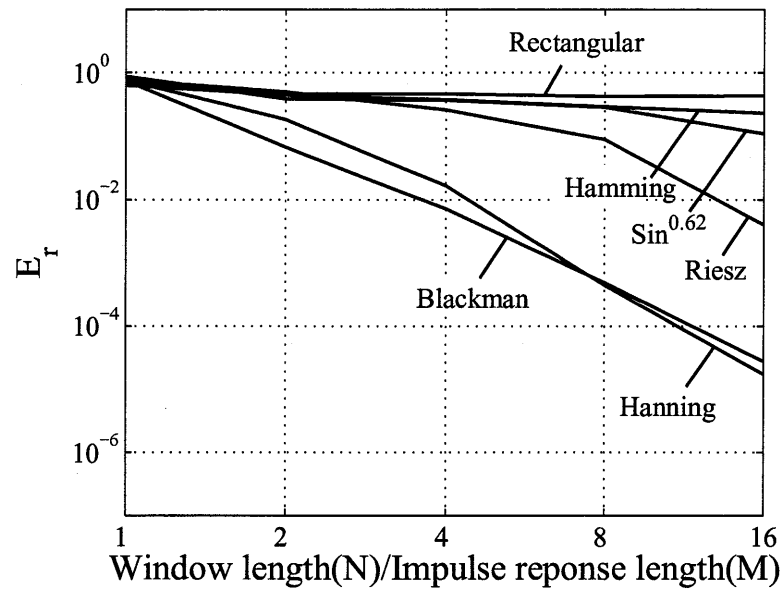


図 4.14 正規化誤差エネルギーと時間窓長 N とインパルス応答長 M の比との関係. 測定用信号はダイナミックレンジ 60 dB の雑音.

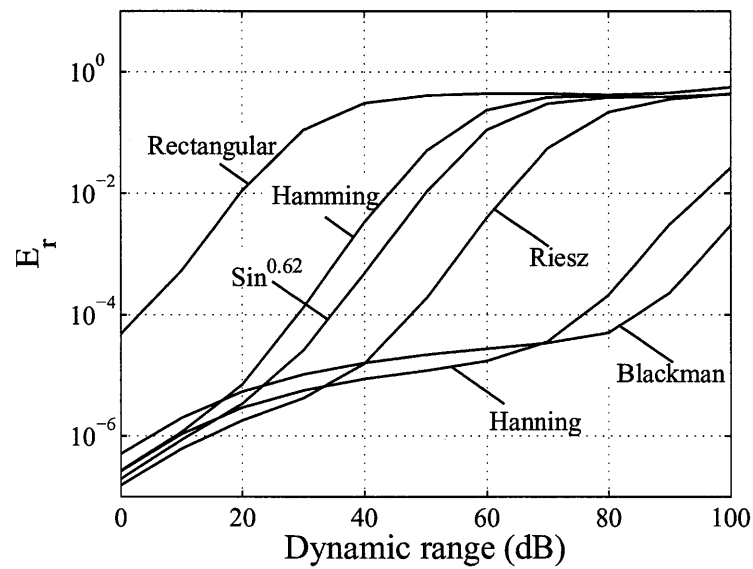


図 4.15 ダイナミックレンジと正規化誤差エネルギーとの関係. $N/M = 16$ について.

第 5 章 暗騒音と高調波歪の影響を軽減 できるインパルス応答測定用信 号の検討

5.1. はじめに

これまで、クロススペクトル法を用いた場合における、音声が入元から任意の観測点までの伝達関数の推定法を示した。この方法では、入元に設置したマイクロフォンで音声を収録し、任意の観測点における音声とのクロススペクトルに基づいて伝達関数を推定している。測定用信号を用いて得られた伝達関数とクロススペクトル法を用いて推定した伝達関数とを比較し、クロススペクトル法を用いた場合においても推定誤差が少ないことを示す必要がある。

この問題に対する解決法として、ダミーヘッドを用いることが考えられる。ダミーヘッドから測定用信号を再生することで測定された伝達関数を真値とする。第 3 章で示したクロススペクトル法と同様の音声を再生・収録を行うことで、ダミーヘッドの入元から観測点までの伝達関数が推定できる。こうして得られた 2 つの伝達関数を比較することで、音声を入力信号としたクロススペクトル法の推定誤差を検証することが出来る。

文献⁵⁶⁾は、人間の音声が大きなパワーを有する帯域において、2 つの伝達関数の差は 1 dB 以下であることを示している。文献⁵⁶⁾で求めた伝達関数は、クロススペクトル法を用いた場合、M 系列を用いた場合について比較している。どちらの条件においても、得られた伝達関数は低域と高域における SNR が低く、SNR の高い伝達関数の比較を行うことは困難であることが指摘されていた。この問題を解決するためには、全帯域について高い SNR で測定が行える測定用信号が必要となる。このような測定用信号は、音響機器や室内伝達関数の高精度測定もできるため応用範囲が広い。

例えば、室内音響分野においてインパルス応答は重要な役割を持っている。残響時間の測定法に関する国際規格においても、インパルス応答に基づくインパルス 2 乗積分法

が規定されている。また、従来はホール等の室内における伝達関数やスピーカ等の音響機器の測定は、長い時間のスイープ音を用いて人間が逐次チェックしながら測定するという非常に手間のかかる方法を用いていた。しかし、これらの特性はインパルス応答から得ることが可能である。そのため、これらの情報を得るためにインパルス応答を測定することが一般的となっている。

従来インパルス応答は、音源信号として継続時間の短いパルスを用いて測定されていた。この方法は最も簡便であるが、暗騒音との SNR を向上させるためには、膨大な回数の同期加算が必要となる。しかし、室内音響特性を時不変システムとみなすことができず、長時間にわたって測定を行うことは、インパルス応答から算出される計算結果に誤差を生じさせる原因となる。

そこで、音源信号としてパルスを用いずに、継続時間の長い測定用の信号を用いてインパルス応答を測定する方法が提案されている。測定用信号には直交する逆信号が存在し、入力信号と収録された信号との相互相関を求めることで、パルスを入力した場合と等価な結果を得ることができる。測定用信号を用いることでパルスのエネルギーを長時間に分散させ、同期加算を行わなくとも SNR の高い測定が可能である。

代表的なインパルス応答測定用信号として、M 系列 (Maximum length sequence)^{67, 76)} , Swept-Sine^{77, 78)} , TSP (Time-Stretched Pulses)^{68, 90)} が挙げられる。これらの信号を用いることで、直接法に比べ高い SNR で測定が可能である。ただし、M 系列を用いた場合、測定されたインパルス応答には高調波歪が多く含まれる。Swept-Sine・TSP を用いて測定されたインパルス応答にも、M 系列より少ないものの高調波歪が含まれる。また、どの信号も全帯域で均一なエネルギーを持つが、実環境における暗騒音のエネルギーは低域に集中するため、低域の SNR が低いという問題点がある。

これらに対して、低域での SNR 改善と高調波歪の除去を目的とした測定用信号である、Logarithmic TSP^{81, 85)} や Logarithmic Swept-Sine^{82, 83)} が提案されている。TSP と Swept-Sine は機能的にほぼ等価なため、本論文では TSP と総称し、線形の TSP を Lin-TSP, Logarithmic TSP を Log-TSP と呼ぶことにする。Log-TSP は、低域にエネルギーが集中しているため、特に低域において測定時の SNR が改善されることを特徴としている。ところが、高域においては Log-TSP を用いて測定した場合の SNR が、Lin-TSP の場合より低くなる。そのため、低域における暗騒音のエネルギーが少ない測定環境では、Log-TSP の SNR は Lin-TSP より低くなる。その場合、従来より用いられている同期加算・信号長の延長・再生音量増加などの SNR 向上手法を使う必要がある。ただし、測定用信号を用いたインパルス応答測定の問題点として、音量増加にともなう測定系の非線

形歪の増大による SNR 向上の限界⁸⁰⁾ や、室内における時変性の影響⁷⁹⁾ が指摘されている。これに対して、暗騒音の振幅周波数特性を用い、位相を適切に設定することにより、SNR を最大化できる TSP が提案されている^{86, 87)}。この TSP は測定環境から振幅周波数特性を設定するため、高調波歪の除去を可能とする制約条件(3.3 節参照)を満たさない場合がありえる。

これらの問題を解決すべく、本研究では、測定環境に適した振幅周波数特性を設定でき、かつ高調波歪を簡便な操作で除去できる測定用信号の提案を行う。本論文では、これらの条件を満たすインパルス応答測定用信号として、Lin-TSP と Log-TSP の特徴を併せ持つ“Warped-TSP”を提案する。まず、Lin-TSP や Log-TSP と Warped-TSP の関係を示し、振幅周波数特性を決めるパラメタを導入する。次に、測定環境で収録した暗騒音から SNR が最大となるパラメタを求める方法について議論する。なお、ここでは、各信号に対する再生系の非線形歪の影響を同程度にするため、単位時間あたりのスピーカに対する入力エネルギーを等しくするという制約条件の下で議論を進める。そして、提案手法を評価するため、暗騒音の特性が異なる 2 つの環境でインパルス応答測定を行い、Lin-TSP・Log-TSP に比べて Warped-TSP が優れていることを示す。

5.2. インパルス応答測定法と問題点

この章では、インパルス応答の測定法について説明する。各測定法の特徴、長所・短所についてまとめる。なお、室内の伝達関数や音響システムは、測定を行う範囲の時間において線形時不変システムとみなし議論を行う。

5.2.1 直接法

最も簡単なインパルス応答測定法は、音源信号としてパルスを用いる方法である。インパルス応答はシステムにパルスを入力した際の出力なので、音源信号としてパルスを用いた場合収録された信号がインパルス応答となる。図 5.1 は直接法におけるインパルス応答測定の流れ図である。入力信号は X 、室内の伝達関数は H であり、インパルス応答は Y となる。直接法の長所は、パルスそのものを放射するため室内の残響を耳で確認できること、そして後処理が必要ないため、測定が簡単なことである。直接法の問題点として、システムに入力できるエネルギーには制約があり、SNR の向上には、測定を繰り返し行い多数の同期加算を必要とすることが挙げられる。この問題点の解決法として、時

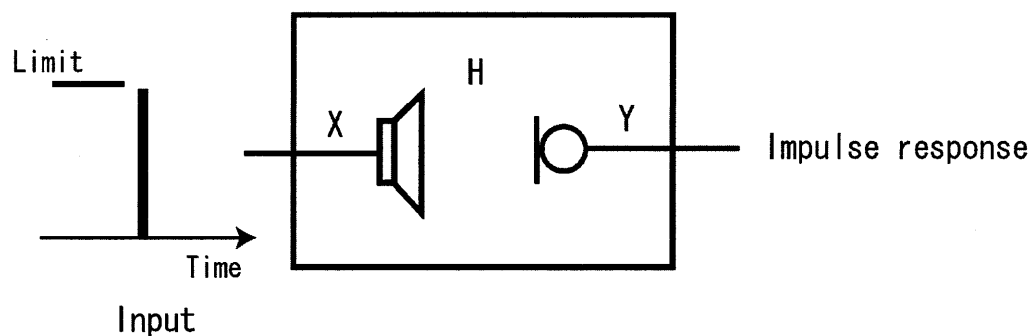


図 5.1 直接法によるインパルス応答測定の流れ図.

間長が短いパルスを入力音源として用いず，パルスのエネルギーを長時間に分散させた信号を入力音源とすることが挙げられる．収録した信号へ後処理を行うことでインパルス応答を求めることができれば，システムに入力できるエネルギーの上限に関する問題点を解決できる．

5.2.2 相互相関法

音源信号としてパルスを用いず，継続時間の長い測定用の信号を用いてインパルス応答を測定する方法がある．直接法は入力できるエネルギーに上限があるため，高い SNR を達成するためには同期加算を行う必要がある．相互相関法では長時間にパルスのエネルギーを分散させることができ，分散した信号の単位時間当たりのエネルギーを向上させ音場に放射することにより，一度の測定で高い SNR を得ることができる (図 5.2)．相互相関法は直交する逆信号 (測定用信号の持つ特性 G に対し， G^{-1}) を持つことが特徴である．収録された測定用信号と逆信号との相互相関関数がインパルス応答となる．

5.2.3 代表的な測定用信号

ホワイトノイズ

ホワイトノイズは自己相関関数がインパルスになる．ホワイトノイズを入力音源として用いた場合，収録された信号と入力音源との相互相関関数がインパルス応答となる．ホワイトノイズは容易に作成することができる信号であるが，自己相関関数を直交させるためには，信号長を十分長くしなければならないという問題点がある．そこで，決めら

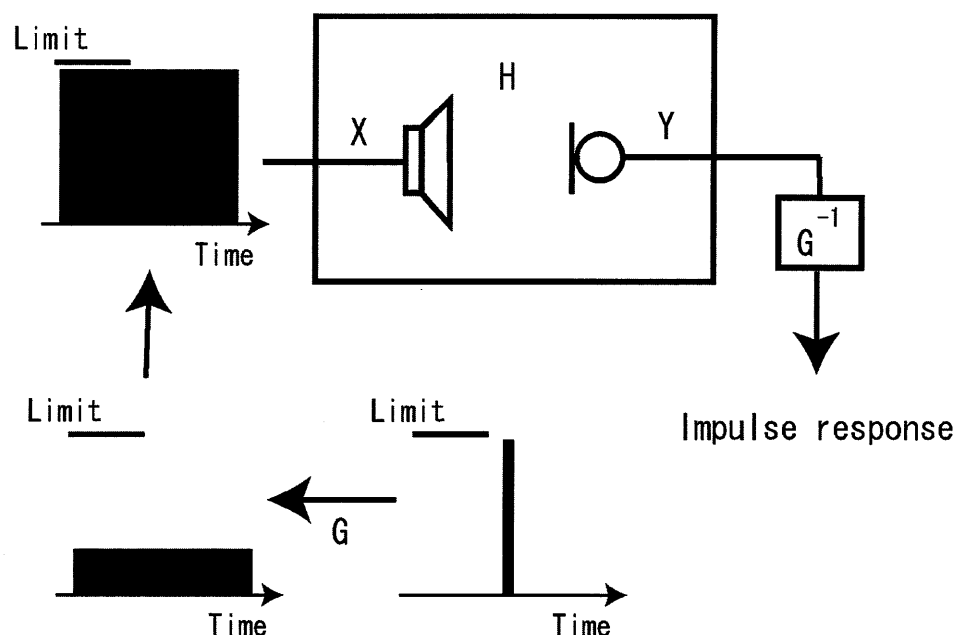


図 5.2 相互相関法によるインパルス応答測定の流れ図.

れた時間長で自己相関関数が直交関係にある疑似ランダムノイズとして，M 系列が提案されている。

M 系列

入力音源に M 系列 (Maximum Length Sequence) を用いることでインパルス応答を測定できる。M 系列は周期的かつ確定的な信号であり，ホワイトノイズのように長い時間の信号を用いなくとも，自己相関関数が完全なインパルスとなる。ただし，M 系列には信号長が $2^n - 1$ 以外で作成することができない制約が存在する。また，M 系列を用いた測定において得られたインパルス応答には，非線形歪を多く含むことが指摘されている。具体的には，非線形歪が測定されたインパルス応答のあらゆる時刻にパルス状となって表れる。

そこで本論文では，時間長への制約が少なく，インパルス応答に含まれる非線形歪みが容易に求められる Lin-TSP と Log-TSP に着目する。

Lin-TSP と Log-TSP

ここでは、Lin-TSP と Log-TSP の特徴と問題点について議論する。ただし、後に行う高調波歪の議論のため、両 TSP 共に低い周波数から高い周波数へ掃引する場合についてのみ考える。高調波歪を分離するためには、低い周波数から高い周波数に掃引する必要があるからである。

Lin-TSP と Log-TSP は、周波数領域の DFT 上で定義されており、逆 DFT により時間波形を求める。Lin-TSP の周波数領域表現 $H_t(k)^{90)}$ を式 (5.1)、Log-TSP の周波数領域表現 $H_l(k)^{81)}$ を式 (5.2) に示す。

$$H_t(k) = \begin{cases} \exp(-j4m\pi k^2/N^2) & (0 \leq k \leq \frac{N}{2}) \\ H_t^*(N-k) & (\frac{N}{2} < k < N) \end{cases} \quad (5.1)$$

$$H_l(k) = \begin{cases} 1 & (k = 0) \\ \frac{\exp(-jak \log(k))}{\sqrt{k}} & (1 \leq k \leq \frac{N}{2}) \\ H_l^*(N-k) & (\frac{N}{2} < k < N) \end{cases} \quad (5.2)$$

$$a = 2m\pi/(N/2)\log(N/2)(m = \text{整数})$$

ここで、 N は信号長、 k は離散周波数番号、 m はパルス幅を決めるパラメタである。また、 $*$ は共役複素を表す記号である。これらをインパルスに戻すための逆信号も容易に求められる^{90, 81)}。

Lin-TSP の振幅周波数特性は全帯域において一定で、Log-TSP の場合は -3 dB/oct となる。従って低域での SNR 向上には Log-TSP を、高域での SNR 向上には Lin-TSP を使った方が良いと考えられる。

5.2.4 相互相関法の特徴と問題点

相互相関法は、直接法において SNR が低いという問題点の解決を図った方法である。長い信号を用いる必要があるというホワイトノイズにおける問題点を M 系列は解決した。

M 系列を用いた場合、信号長に制約があることと、測定されたインパルス応答に含まれる高調波歪が問題点となる。Lin-TSP は、M 系列が持つ時間長における制約を解決した。Log-TSP は、Lin-TSP と同様に時間長における制約が無く、実環境における暗騒音の振幅周波数特性が低域に集中している特徴に着目して提案された測定用信号である。また、測定されたインパルス応答に含まれる高調波歪も分離可能であるという特徴を有している。高調波歪の分離に関する議論は次章にて行う。Log-TSP は、測定環境の暗騒音におけるエネルギーが低域に偏っている環境において、Lin-TSP より高い SNR でインパルス応答を測定できる。

だが、測定環境における暗騒音は必ずしも Log-TSP に有利であるとは限らない。例えば無響室等においては低域の暗騒音も十分少なく、平坦な振幅周波数特性であるため Log-TSP より Lin-TSP のほうが高い SNR となる。このように、相互相関法は直接法に比べ SNR は高いものの、さらに高い SNR を達成するためには、測定環境における暗騒音の特徴に応じて測定用信号を選択しなくてはならない。これは、暗騒音の振幅周波数特性は測定環境により異なる一方、測定用信号の振幅周波数特性は固定であることに問題がある。そのため、測定環境に応じて振幅周波数特性を変えるパラメタを持ち、Log-TSP が有する高調波歪の分離という特徴を備えた測定用信号が存在すれば、これらの問題点を解決できる可能性がある。

本論文では、Log-TSP では低域の SNR が高く、Lin-TSP では高域の SNR が高いという特徴に注目し、両信号の長所を兼ね備えた測定用信号を提案することでこの問題点の解決を図る。具体的には、低域側が Log-TSP、高域側が Lin-TSP となるよう両信号を接続した新たな測定用信号“Warped-TSP”を提案する。Warped-TSP は、Log-TSP と Lin-TSP を任意の周波数で接続するパラメタを持ち、このパラメタを操作することで振幅周波数特性を操作できる。このパラメタを測定環境に応じて変化させることにより、従来から用いられている測定用信号以上の SNR でインパルス応答の測定が可能となる。また、Warped-TSP は Log-TSP が持つ高調波歪が分離可能である特徴も有している。測定環境に適したパラメタの設定方法については 5.3.5 節で述べる。

5.3. Warped-TSP の定義と特徴

これまで、相互相関法によるインパルス応答測定法について述べ、特に Lin-TSP、Log-TSP の定義式と特徴・問題点について説明した。本章では、Lin-TSP、Log-TSP の長所を併せ持つ Warped-TSP の提案を行い、特徴を説明する。

5.3.1 定義式

まず、提案する Warped-TSP を DFT 上で以下のように定義する.

$$H(k) = \begin{cases} 1 & (k = 0) \\ b(k) \exp\{-jC(w(k)\theta_l(k) \\ + (1 - w(k))\theta_h(k))\} & (1 \leq k \leq \frac{N}{2}) \\ H^*(N - k) & (\frac{N}{2} < k < N) \end{cases} \quad (5.3)$$

$$\begin{aligned} \theta_l(k) &= a_2 k \log(k) / \beta \\ \theta_h(k) &= \frac{1}{\beta} \left\{ \frac{a_1}{2} k^2 + \left(2\pi - a_1 \frac{N}{2} \right) k + a_3 \right\} \\ w(k) &= 1 / \{1 + \exp(\alpha(k - n_t))\} \\ b(k) &= w(k) \frac{1}{\sqrt{k}} + (1 - w(k)) \frac{1}{\sqrt{n_t}} \\ n_t &= \text{round}(N f_c / f_s) \end{aligned}$$

ここで、 $\theta_l(k)$ は Log-TSP の特徴を、 $\theta_h(k)$ は Lin-TSP の特徴を持っている. この2つの関数は、接続周波数 f_c において、シグモイド関数を用いた混合関数 $w(k)$ によって滑らかに接続されている. したがって、接続周波数 f_c が 0 に近ければ Lin-TSP に近い特徴を持ち、 $f_s/2$ に近ければ Log-TSP に近い特徴を持つ. Warped-TSP は、このように周波数軸の伸縮特性が接続周波数付近で変化することを特徴とする. $\alpha(> 0)$ は接続部分の帯域幅を決定するパラメタであり、0 に近いほど帯域幅が広がる. N は信号長、 k は離散周波数番号、 β は信号長と掃引区間長の比であり、Lin-TSP・Log-TSP における m の定義を $m = N/2\beta$ とした場合と等価である. 信号長が N ならば、掃引区間は N/β となる. $b(k)$ は振幅周波数特性を表す関数である. 係数 a_1 , a_2 , a_3 , C はこれらのパラメタを用いて以下のように表すことができる.

$$\begin{aligned} a_1 &= \frac{2\pi}{n_t \log(n_t) + N/2} \\ a_2 &= \frac{a_1 n_t + 2\pi - a_1 N/2}{1 + \log(n_t)} \\ a_3 &= a_2 n_t \log(n_t) - \frac{a_1 n_t^2}{2} - \left(2\pi - \frac{a_1 N}{2} \right) n_t \end{aligned}$$

$$C = \frac{C_1 - \text{mod}(C_1, \pi)}{C_1}$$

$$C_1 = w(N/2)\theta_l(N/2) + (1 - w(N/2))\theta_h(N/2)$$

逆信号も DFT 上で定義でき、以下のようになる.

$$H(k)^{-1} = \begin{cases} 1 & (k = 0) \\ \frac{1}{b(k)} \exp\{jC(w(k)\theta_l(k) \\ + (1 - w(k))\theta_h(k))\} & (1 \leq k \leq \frac{N}{2}) \\ H^{-1*}(N - k) & (\frac{N}{2} < k < N) \end{cases} \quad (5.4)$$

5.3.2 Warped-TSP の性質

ここでは、Warped-TSP の基本的な性質について述べる. 図 5.3 は、Warped-TSP の波形 (左上) と振幅周波数特性 (右上), 逆 Warped-TSP の波形 (左下) と振幅周波数特性 (右下) を示している. ここでは、 $N = 65536$, $\beta = 2$, $\alpha = 0.1$ とし、接続周波数 f_c を 3000 Hz, サンプル周波数 f_s を 44.1 kHz としている. Warped-TSP は、接続周波数まで Log-TSP と同様に -3dB/oct の減少特性を持ち、接続周波数以上では Lin-TSP と同様に一定の振幅周波数特性を持つ.

図 5.4 に接続周波数を 125Hz から 8000Hz まで 1 オクターブ毎に設定し作成した Warped-TSP の振幅周波数特性を示す. N , α , β は、それぞれ 65536, 0.1, 2 であり、 f_s を 44.1 kHz である. また、作成した全ての信号のエネルギーを等しくしている. この図からも Warped-TSP は、Log-TSP と Lin-TSP を滑らかに接続したものであることが分かる.

5.3.3 高調波歪の影響

Lin-TSP を用いた場合のインパルス応答における高調波歪は、高い周波数から低い周波数へ掃引する信号として現れ、低域において主応答と極めて近くなるため除去が難しい. これに対して、Log-TSP では主応答の手前にパルスとなって現れる⁸⁴⁾ ため、主応答の手前の応答を取り除くという簡便な操作で、高調波歪を除去できる. 高調波歪が除去

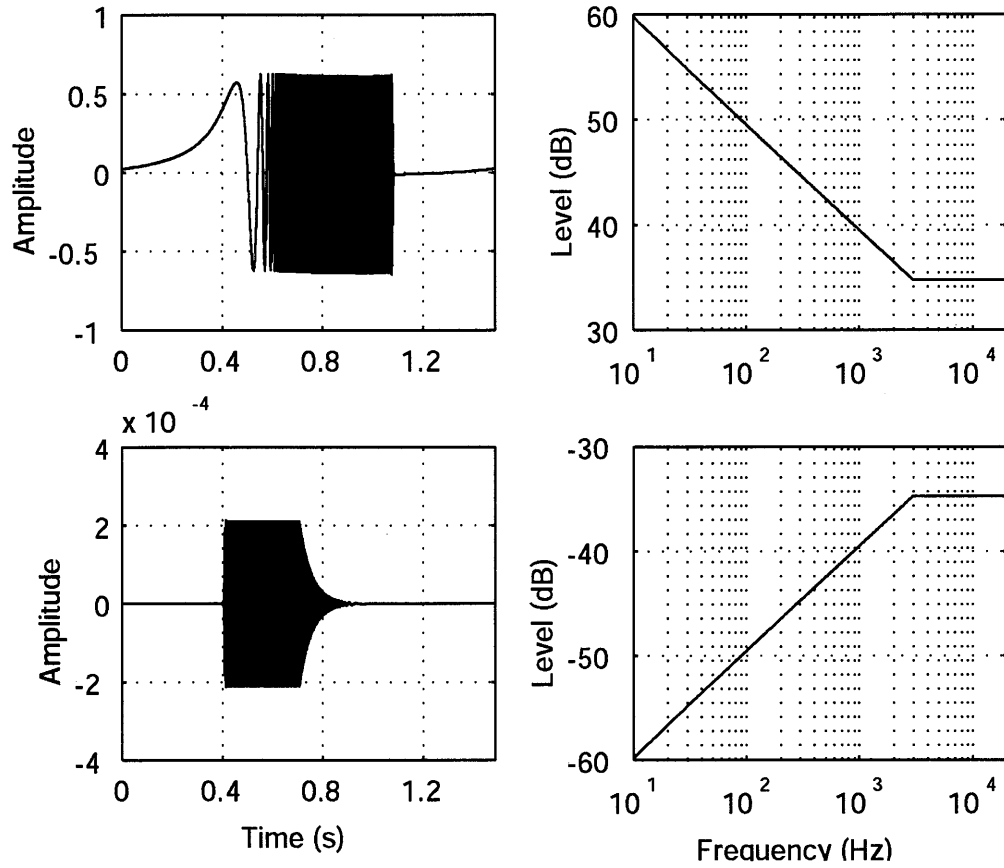


図 5.3 Warped-TSP(上図) と逆 Warped-TSP(下図) の時間波形 (左図) と振幅周波数特性 (右図).

できるためには、インパルス応答に含まれる高調波歪の群遅延特性が全ての周波数において主応答の手前の時刻に表れるという制約条件を満たす必要がある。

ここでは、Warped-TSP を用いて測定されたインパルス応答における高調波歪の影響を調べるために、Warped-TSP における高調波の群遅延特性について議論する。Warped-TSP は、接続周波数以下では Log-TSP、接続周波数以上では Lin-TSP の特徴を持つ。そこで、接続周波数以下と以上に分けて考えた場合の Warped-TSP 信号自体の群遅延特性は、Warped-TSP における位相特性 $\theta_l(k)$ と $\theta_h(k)$ の負の導関数であり

$$\begin{aligned}\tau_l(k) &= -\frac{d\theta_l(k)}{dk} \\ &= \frac{Ca_2}{\beta} (1 + \log(k)) \frac{N}{2\pi f_s}\end{aligned}\quad (5.5)$$

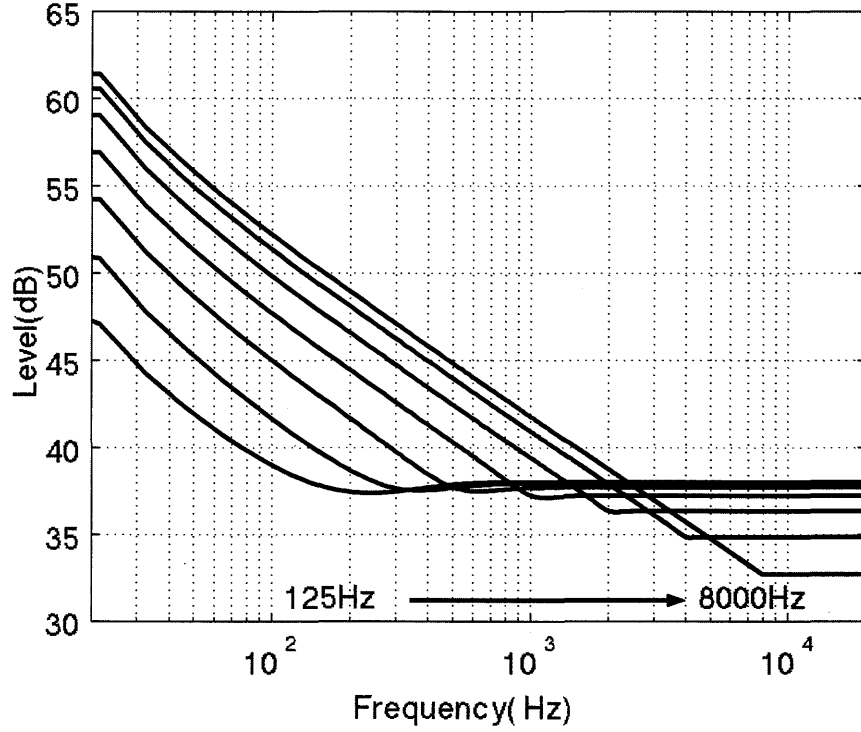


図 5.4 Warped-TSP における，接続周波数による振幅周波数特性の変化．接続周波数は，125 Hz から 8000 Hz までオクターブ毎に設定している．また，各信号が等しいエネルギーを持つようにしている．

$$\begin{aligned}
 \tau_h(k) &= -\frac{d\theta_h(k)}{dk} \\
 &= \frac{C}{\beta} \left\{ a_1 k + 2\pi - a_1 \frac{N}{2} \right\} \frac{N}{2\pi f_s}
 \end{aligned} \tag{5.6}$$

で表される．これらに対応する第 m 次高調波の群遅延特性は

$$\begin{aligned}
 \tau_{m,l}(k) &= \frac{Ca_2}{\beta} \left\{ 1 + \log \left(\frac{k}{m} \right) \right\} \frac{N}{2\pi f_s} \\
 &= \tau_l(k) - \frac{Ca_2}{\beta} \frac{N}{2\pi f_s} \log(m)
 \end{aligned} \tag{5.7}$$

$$\tau_{m,h}(k) = \frac{C}{\beta} \left\{ a_1 \frac{k}{m} + 2\pi - a_1 \frac{N}{2} \right\} \frac{N}{2\pi f_s} \tag{5.8}$$

となる．群遅延領域において，畳み込みは和で表されるため，インパルス応答に含まれる高調波歪は，これら高調波の群遅延特性に逆信号を加算した時刻に現れることになる．

逆 Warped-TSP の群遅延特性は、Warped-TSP の群遅延特性の符号を反転したものである。インパルス応答に含まれる第 m 次高調波歪の群遅延特性は、

$$\begin{aligned}\tilde{\tau}_{m,l}(k) &= \tau_{m,l}(k) - \tau_l(k) \\ &= -\frac{Ca_2}{\beta} \frac{N}{2\pi f_s} \log(m)\end{aligned}\tag{5.9}$$

$$\begin{aligned}\tilde{\tau}_{m,h}(k) &= \tau_{m,h}(k) - \tau_h(k) \\ &= -\frac{C}{\beta} \frac{N}{2\pi f_s} \frac{m-1}{m} a_1 k\end{aligned}\tag{5.10}$$

となる。

Warped-TSP を用いた場合の第2次・第3次高調波歪の群遅延特性を図5.5に示す。 m 次高調波歪は接続周波数以下において、群遅延の値が定数値、すなわちパルスとなる。また、接続周波数以上では、高い周波数から低い周波数へ掃引する信号となる。次数が高くなるに従い $\tilde{\tau}_{m,l}$ と $\tilde{\tau}_{m,h}$ が交わる周波数は高くなるが、Warped-TSP を用いた測定において、 m 次高調波が主応答（太線）に重なることは無く、主応答の手前に高調波歪が現れることになる。そこで、主応答以前の応答を取り除くことより、Warped-TSP は Log-TSP の場合と同様に高調波歪を除去できる。以下の実験では、低い周波数から高い周波数への掃引を用い、インパルス応答の主応答より前の部分を除去している。

5.3.4 直線状畳み込みにおける環状シフト量

設計長より長いインパルス応答を測定する場合、直線状畳み込みを行う必要がある。このとき、振幅が十分収束していない時刻を原点とすると、畳み込み演算における誤差が大幅に増加する。これを防ぐために、測定用信号を環状にシフトすることで誤差を低減させる手法が一般に用いられている。Lin-TSP に関しては、信号長 N とパルス幅 m からシフト量の最適値を定めることができる^{88, 89, 90}。ところが、Warped-TSP におけるシフト量の最適値は、信号長 N 、掃引区間比率 β 、接続周波数 f_c から一意に定めることができない。そのため、本論文における測定では、掃引区間外において最も振幅が0に近い時刻が原点となるように環状シフト量を定める。

図5.6は、次節で実験に用いた Warped-TSP の波形である。上図は定義式から作成された Warped-TSP であり、下図が適切な環状シフトを行ったものである。開始点と終了点における振幅が0ではない場合、音の開始と終了に雑音が含まれる。この雑音が原因

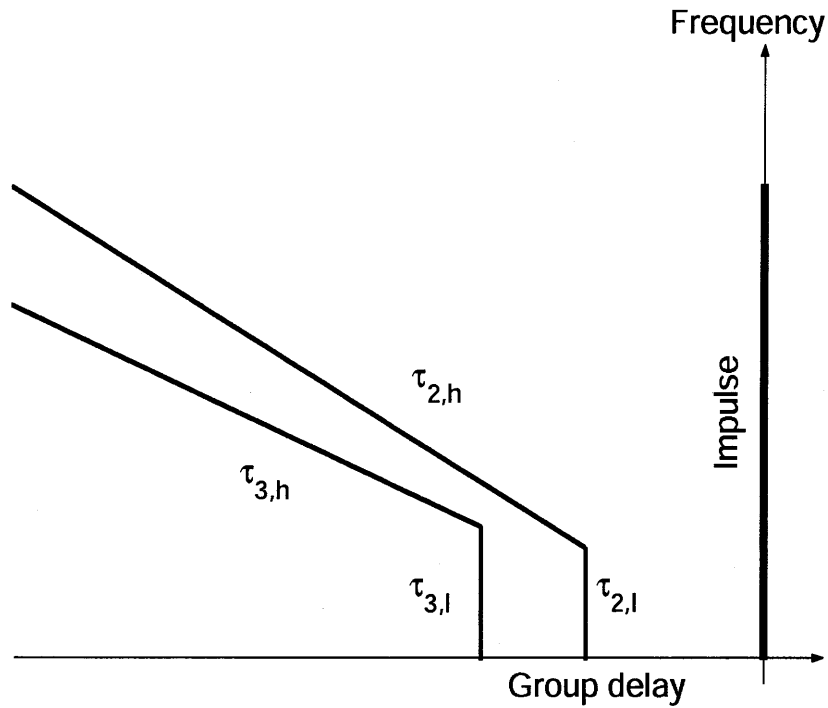


図 5.5 Warped-TSP の第 2 次・第 3 次高調波歪 ($\tau_{2,h}$ and $\tau_{2,l}$, $\tau_{3,h}$ and $\tau_{3,l}$). 高調波歪は太線の垂線で示される主応答とは時間的に重ならない。

で、測定されたインパルス応答の SNR が低くなる。環状シフトにより振幅が 0 に近い地点を原点とすることにより、この問題を解決することができる。

5.3.5 測定環境に適したパラメタの設定法

Warped-TSP の接続周波数は、測定環境毎に設定する必要がある。ここでは、測定環境の暗騒音を入力とし、最も SNR が高くなる接続周波数を求める方法を提案する。

インパルス応答に含まれる暗騒音のエネルギーは、暗騒音のみを収録し逆信号を畳み込んだ信号のエネルギーと等価である。そのため、雑音低減比 (Noise Reduction Ratio, NRR) を、以下の式のように、収録した暗騒音 $n(n)$ のエネルギーと、暗騒音に逆信号を畳み込んだ信号のエネルギーとの比とする。

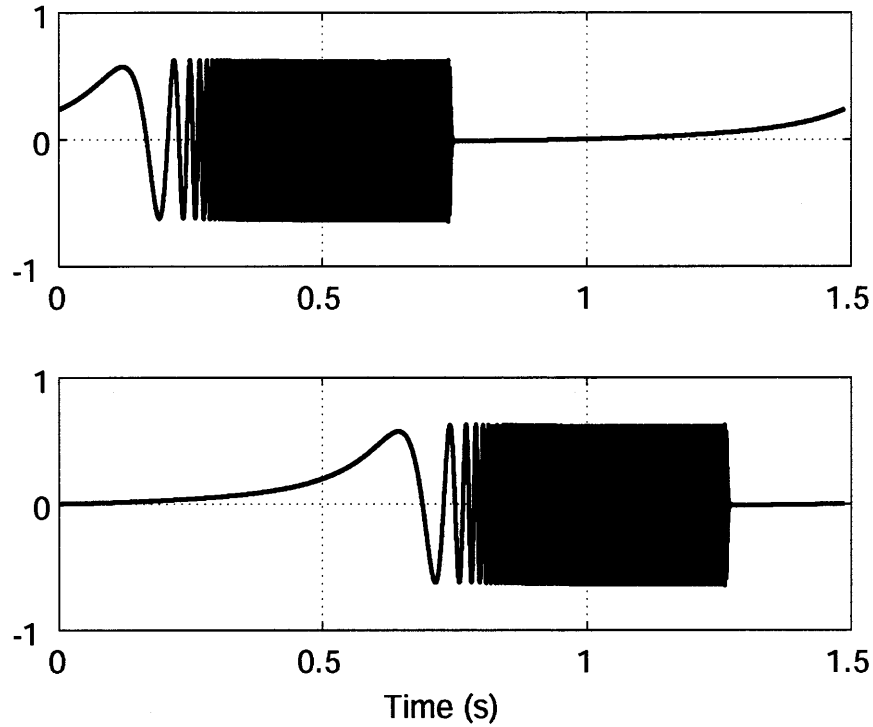


図 5.6 定義式に基づいて作成した Warped-TSP(上図) と, 適切な環状シフトを行った Warped-TSP(下図).

$$NRR = 10 \log_{10} \left(\frac{\sum_{n=0}^{N-1} n(n)^2}{\sum_{n=0}^{N-1} (n(n) * h^{-1}(n))^2} \right) \quad (5.11)$$

N は, 信号長である. また $*$ は畳み込みを表している.

この NRR の値が最も高くなるように f_c を設定すれば, 測定環境に適した Warped-TSP が得られる. なお, 逆信号は, 測定用信号と畳み込んだときに得られるインパルスの振幅が 1 となるように決定する. ここで各測定用信号は, 再生系で生じる非線形歪への影響をほぼ一定とするため, エネルギーが等しくなるように設計する. 従って, 逆信号のエネルギーは測定用信号に依存することに注意する.

図 5.7 に, 防音室を含む様々な環境で測定した暗騒音の 1/3 オクターブバンド毎のエネルギーを示す. 暗騒音の振幅周波数特性は環境により異なっており, 多くの環境において, エネルギーは低域に集中していることが分かる. だが, 一部の環境は比較的平坦な

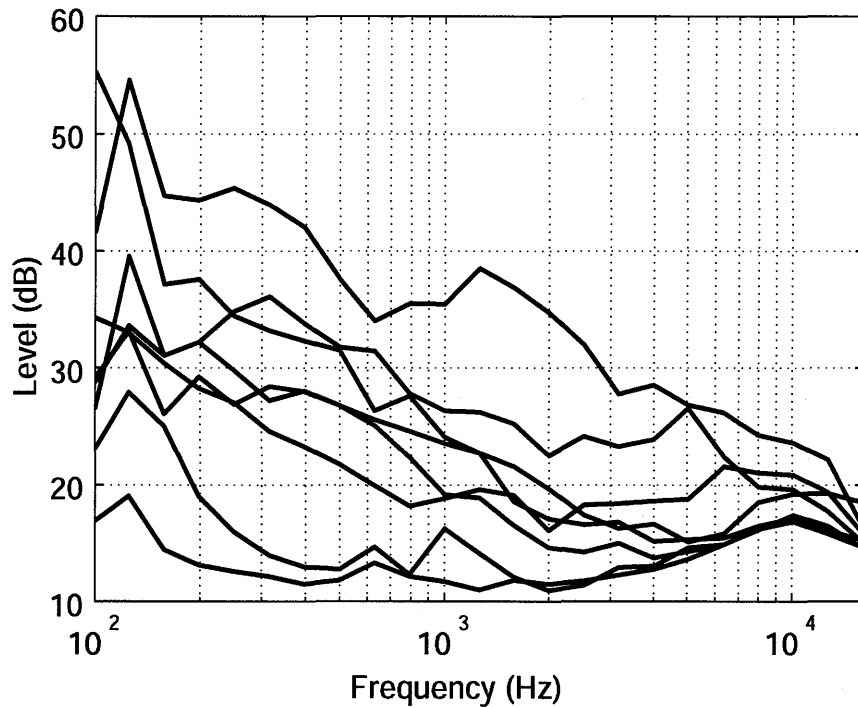


図 5.7 様々な場所で測定した暗騒音の 1/3 オクターブバンド毎のエネルギー。

振幅周波数特性を持つことも分かる。Warped-TSP は接続周波数を操作することにより、低域にエネルギーを集中させることも、平坦な振幅周波数特性を持たせることも可能である。このことから Warped-TSP は、実環境におけるインパルス応答測定に適しているといえる。

5.3.6 Warped-TSP のまとめ

各 TSP の特徴を表 5.1 にまとめる。Lin-TSP, Log-TSP は信号長と全体の長さに対するパルス幅をパラメタとして持っているが、振幅周波数特性を操作する自由度を持っていない。そのため、測定環境に応じて信号を使い分ける必要がある。提案する Warped-TSP は、異なる振幅周波数特性を持つ環境においても、暗騒音に基づいて接続周波数を設定できるパラメタを持つ。適切なパラメタを設定することにより、Lin-TSP・Log-TSP より SNR の高い測定を行うことができる。

表 5.1 各 TSP の特徴

信号の種類	linear TSP	Log-TSP	Warped-TSP
低域の SNR	低	高	中
高域の SNR	高	低	中
適した環境	防音室	実環境	環境に適応
高調波歪	除去不能	除去可能	除去可能

5.4. 実環境における実験

実環境でインパルス応答の測定実験を行い，Warped-TSP を評価する．特に測定環境に応じたパラメタ設定が，実環境における測定で有効で働くことを示す．また，高調波歪が分離可能であるかについても測定実験を行う．

5.4.1 収録条件

Warped-TSP の有効性を確認するために，実環境で比較実験を行った．実験は，図 5.8, 5.9 に示す条件で，異なる環境においてダミーヘッドを用いたヘッドフォン測定とスピーカ測定を行った．測定されたインパルス応答とインパルス応答に含まれる暗騒音のエネルギーとの SNR を指標とし，評価を行った．ヘッドフォンの測定環境は簡易防音室，スピーカの測定環境は防音室である．以下では，ヘッドフォン測定を行った環境を測定環境 1，スピーカ測定を行った環境を測定環境 2 とする．比較には Lin-TSP と Log-TSP を用いた．AD/DA 変換機には，両測定において M-Audio のオーディオインタフェース FireWire 410 を使用し，直接コンピュータに音を取り込んだ．サンプリング周波数は 44.1 kHz，量子化ビット数は 24 ビットとした．また，全ての測定用信号を公平に評価するため，信号長を等しくし，エネルギーが等しくなるようにした．信号長 N を 65536 とし，Lin-TSP, Log-TSP におけるパルス幅 $m = N/4$ とした．Warped-TSP では $\alpha = 0.1$, $\beta = 2$ とした．また，各測定用信号で同期加算回数を 10 回とした．直線状畳み込みにおける影響を除去するため各測定用信号を 12 個連結し，環状畳み込みで計算されたインパルス応答の最初と最後を除いた 10 回を加算した．さらに，各測定用信号を 1 つだけ使い，直線状畳み込みによって計算された場合との比較も行った．

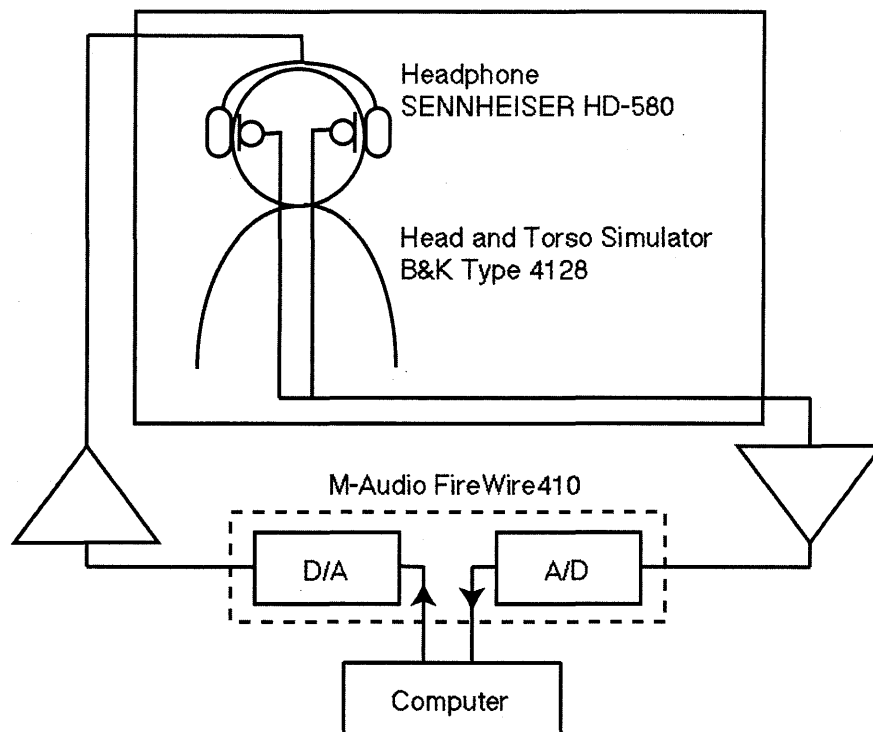


図 5.8 測定環境 1 における収録条件 (簡易防音室・ヘッドホン測定)

5.4.2 測定環境に適した接続周波数の設定

まず, 図 5.8, 5.9 の環境において暗騒音を 3 秒程度収録した. 次に接続周波数を 1/3 オクターブの中心周波数に設定した逆 Warped-TSP と収録した暗騒音から雑音低減比 NRR の値を式 (5.11) によって求めた. この時の接続周波数と雑音低減比の関係を図 5.10 に示す. この図より, 適切な接続周波数は測定環境 1 (簡易防音室・ヘッドホン測定) では 3150 Hz, 測定環境 2 (防音室・スピーカ測定) では 250 Hz となることが分かる. 測定に用いた TSP に関する条件を表 5.2, 5.3 にまとめる.

5.4.3 測定用信号の評価指標

測定されたインパルス応答には, 暗騒音の影響が含まれる. このインパルス応答を室内残響時間よりも十分長く切り出すと (例えば数倍) 後半部分は暗騒音成分のみと見なせる. そこで, 図 5.11 のようにインパルス応答長を N とし, 暗騒音が定常的と仮定すると, 後半部分 ($N/2$ から $N - 1$ まで) のエネルギーを 2 倍することにより, 全体の暗騒音レベ

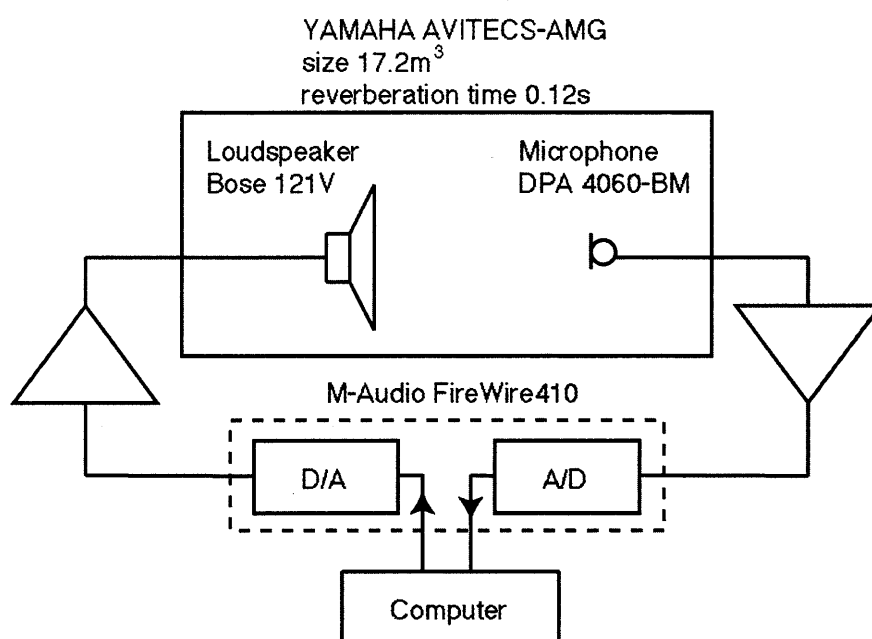


図 5.9 測定環境 2 における収録条件 (防音室・スピーカー測定)

表 5.2 測定条件 (測定環境 1)

信号の種類	linear TSP	Log-TSP	Warped-TSP
サンプリング	44.1kHz, 24bit		
信号長	65536		
パルス幅	32768		
接続周波数の選択	-	-	3150Hz

ルを計算できる。そのため、インパルス応答の SNR は、次式で計算できる。

$$SNR = 10 \log_{10} \left(\frac{\sum_{n=0}^{N-1} x(n)^2 - 2 \sum_{n=N/2}^{N-1} x(n)^2}{2 \sum_{n=N/2}^{N-1} x(n)^2} \right) \quad (5.12)$$

表 5.3 測定条件 (測定環境 2)

信号の種類	linear TSP	Log-TSP	Warped-TSP
サンプリング	44.1kHz,24bit		
信号長	65536		
パルス幅	32768		
接続周波数の選択	-	-	250Hz

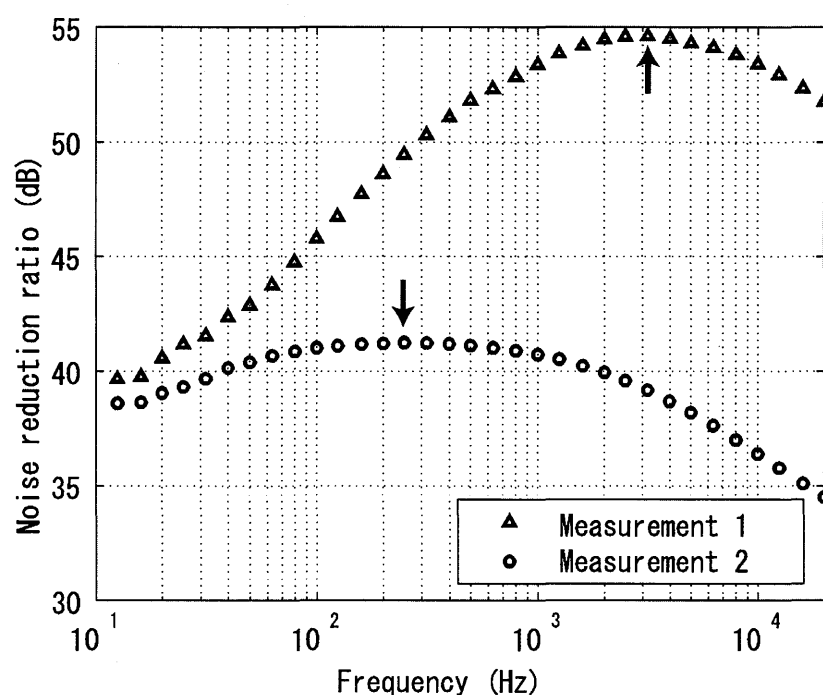


図 5.10 各測定環境における接続周波数と雑音低減比の関係および適切な接続周波数の設定

5.4.4 評価結果

始めに測定環境の暗騒音と、測定に用いた各測定用信号の振幅周波数特性を比較する。図 5.12 は測定環境 1 について、図 5.13 に測定環境 2 についてであり、上図は暗騒音、下図は各測定用信号の 1/3 オクターブバンド毎のエネルギーを示している。図 5.8 の測定環境 1(簡易防音室・ヘッドホン測定)の暗騒音は、低域にエネルギーが集中し、図 5.9 の測定環境 2(防音室・スピーカー測定)の暗騒音は、低域と高域のエネルギーが同程度となって

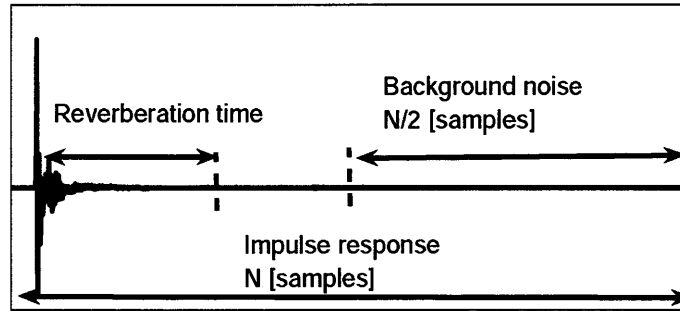


図 5.11 暗騒音区間の定義.

いる. Lin-TSP や Log-TSP は, 両測定環境において同じ信号で測定を行っているが, 提案する Warped-TSP は, 先の 3.1 節で求めた適切な接続周波数を用いて設計した. すなわち, 低域のエネルギーが大きい測定環境 1 では適切な接続周波数が 3150 Hz となるため Log-TSP に近くなり, 低域と高域が同程度の測定環境 2 では適切な接続周波数が 250 Hz であるため, Lin-TSP に近くなる. また, 測定に使用した各 TSP のスペクトログラムを図 5.14, 5.15, 5.16, 5.17 に示す. 測定環境 1 における Warped-TSP は低域の掃引時間が長く, 測定環境 2 における Warped-TSP はほぼ Lin-TSP と等しいスペクトログラムを持つことが分かる.

インパルス応答の SNR による評価結果を図 5.18 に示す. 同期加算した場合と, 同期加算を行わず各インパルス応答より SNR を計算し, 平均したものを比較している. 環境, 測定用信号の種類に依存せず, 10 回の同期加算によって SNR がほぼ 10 dB 改善している. これは, 測定環境における暗騒音に概ね相関が無いことを示唆している. また, 測定環境 1 における暗騒音 (図 5.12(a)) は低域にエネルギーが偏っているため, Lin-TSP の場合の SNR が最も低い. 逆に, 測定環境 2 における暗騒音 (図 5.13(b)) は低域と高域でのエネルギーが同程度であるため, 高域で SNR が低くなる Log-TSP の場合の SNR が, 全体としては最も低くなる. これに対し, 提案する Warped-TSP では, 測定環境に応じて適切な接続周波数を設定することができるため, Lin-TSP ・ Log-TSP より SNR の高い測定を行えることが分かる.

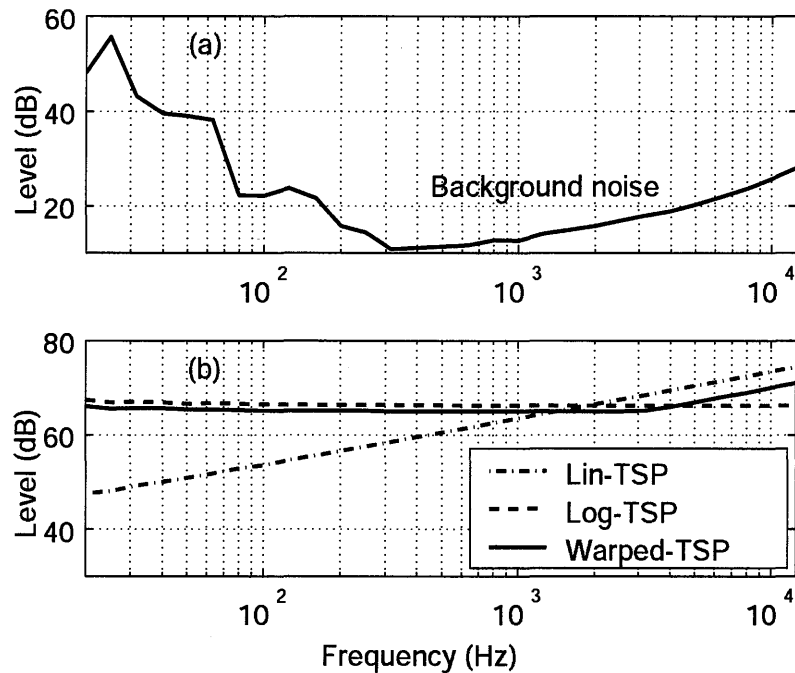


図 5.12 測定環境 1(図 5.8) の暗騒音 (a) と、各測定用信号 (b) の、1/3 オクターブバンド毎のエネルギー。

5.4.5 直線状畳み込みについての検討

前節と同様の実験を直線状畳み込みによって行った。図 5.19 に、環状畳み込みと直線状畳み込みにおける SNR を示す。全ての測定において、環状畳み込みを用いた場合と直線状畳み込みを用いた場合における SNR は、ほぼ等しいことが分かる。また、環状畳み込みの場合と同様に、10 回の同期加算によって、ほぼ 10 dB SNR が改善していた。このことより直線畳み込みを用いても問題なく使用できることが分かる。

5.4.6 高調波歪の除去に関する実験

本章の 3.3 節において、各 TSP で測定されたインパルス応答に含まれる高調波歪について示し、Log-TSP, Warped-TSP については高調波歪が除去可能であることを示した。本節では、各 TSP で発生する高調波歪の測定を行い、高調波歪が実環境における測定において除去可能であるか考察を行う。

測定環境は無響室を利用した。信号長は 65536, サンプル周波数は 44.1 kHz, 量子

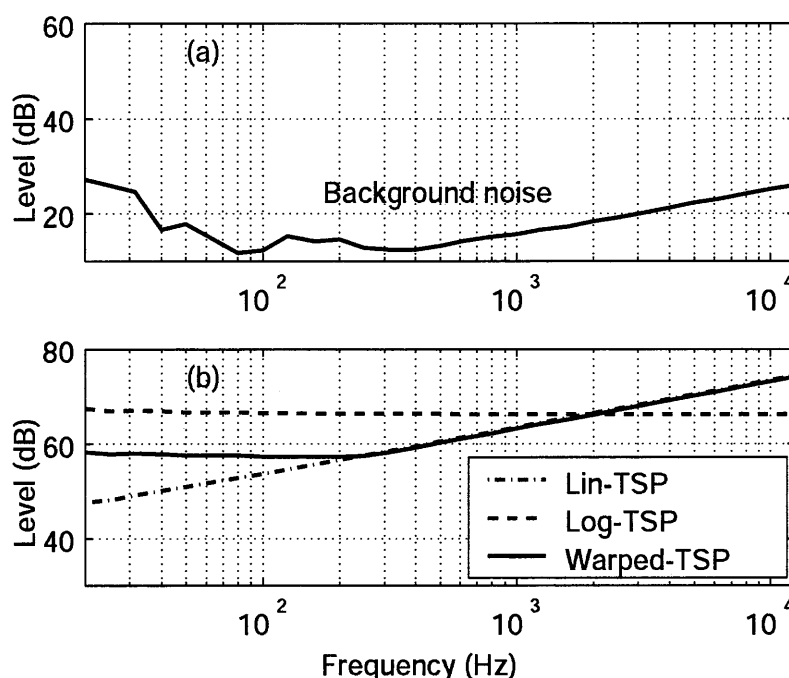


図 5.13 測定環境 2(図 5.9) の暗騒音 (a) と、各測定用信号 (b) の、1/3 オクターブバンド毎のエネルギー。

化ビット数は 24Bit とし、収録に用いた装置は 4.1 節と同様にした。Warped-TSP に関しては、暗騒音に基づいた接続周波数を設定せず、Lin-TSP 部分と Log-TSP 部分が等しい時間となるように接続周波数を設定した。スピーカとマイクロフォンの間隔を 50cm に設定し、入力音圧を高く設定した。具体的には、Lin-TSP を精密騒音系で観測し、最大値が 90 dB を示す程度とした。これは、高調波歪が入力音圧に比例して増大する特徴を持ち、日常的に用いる音圧レベルでは高調波歪のエネルギーより暗騒音のエネルギーが支配的になるためである。

測定されたインパルス応答のスペクトログラムを図 5.20, 5.21, 5.22 に示す。高調波歪は主応答より手前の時刻に観測されるため、主応答より 0.4 秒程度手前を原点として切り出した。中央付近に観測される縦の直線が主応答である。また、主応答より手前の時刻に観測される線が高調波歪である。Lin-TSP では、高い周波数から低い周波数に掃引する直線として観測され、低域において主応答と重なっている。Log-TSP における高調波歪は、全ての周波数で等しい時刻に観測されており、低域においても主応答と重ならない。そのため、主応答より手前を切り捨てることで除去可能である。Warped-TSP でも、

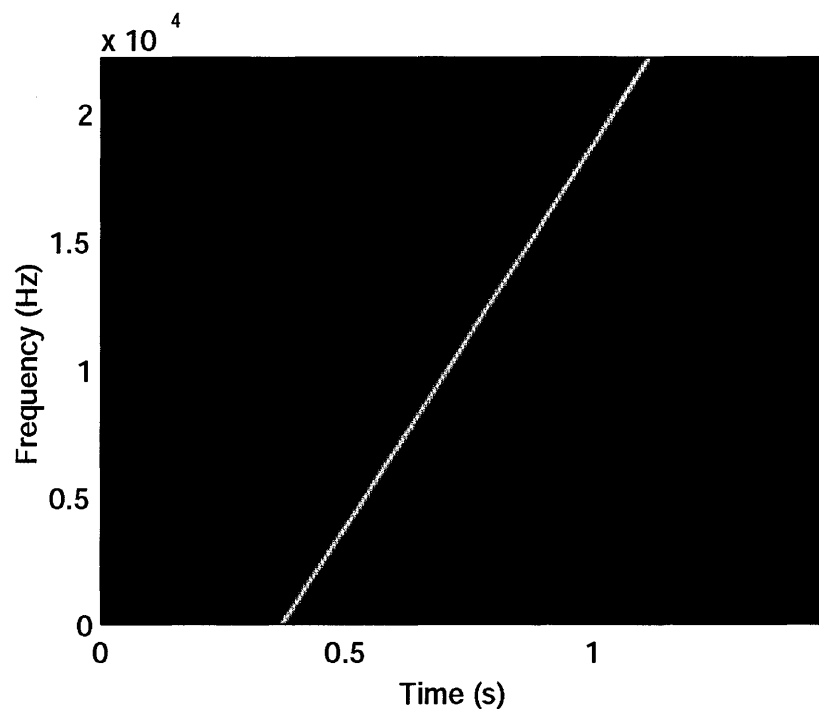


図 5.14 測定に用いた Lin-TSP のスペクトログラム

Log-TSP と同様にある周波数以下の高調波歪は等しい時刻に観測される．また，ある周波数以上では Lin-TSP と同様に，高い周波数から低い周波数へ掃引する直線として観測されている．このことから，Warped-TSP における高調波歪は，Log-TSP と同様に主応答と分離していることが分かる．したがって，主応答より手前の時刻の応答を切り捨てることで高調波歪を除去することが可能である．

5.5. まとめ

本論文では，Lin-TSP，Log-TSP を組み合わせることにより，いずれの方法よりも高い SNR で測定が可能な Warped-TSP の設計法について述べた．始めに，Warped-TSP の設計式・特徴を述べ，Warped-TSP が Log-TSP と同様に，高調波歪が除去できることを示した．また，測定環境での暗騒音を入力とし，接続周波数のパラメタを操作することで，測定環境に適した Warped-TSP が作成できることを示した．次に，インパルス応答の測定を防音室と簡易防音室で行い，測定されたインパルス応答と，インパルス応答に含まれる暗騒音のエネルギーとの SNR を指標とした評価を行った．その結果，Warped-TSP

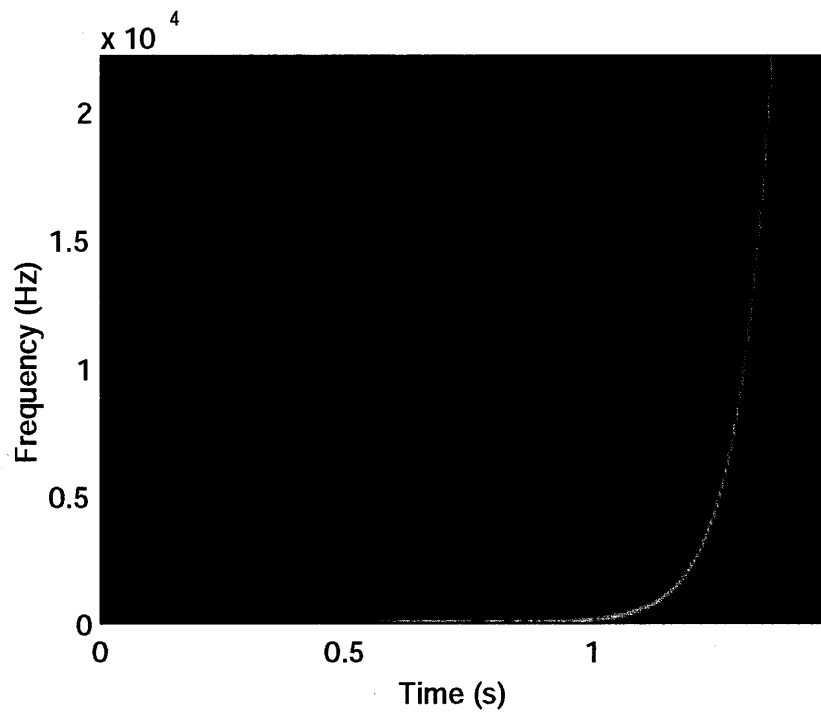


図 5.15 測定に用いた Log-TSP のスペクトログラム

は、測定環境に適したパラメタを設定することで、Lin-TSP・Log-TSP より高い SNR で測定できることが示された。今後は、高調波歪が除去可能な特徴を持ちつつ、暗騒音の振幅周波数特性の近似精度を高めたインパルス応答測定用信号について検討を行う予定である。また、測定系の振幅周波数特性を考慮した接続周波数の設定法についても検討を行う。

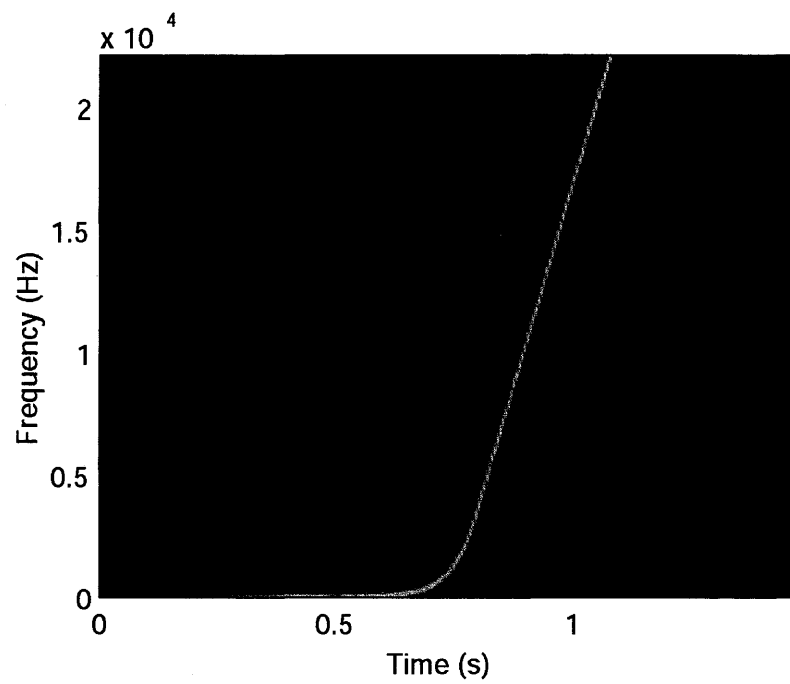


図 5.16 測定に用いた Warped-TSP のスペクトログラム (測定環境 1)

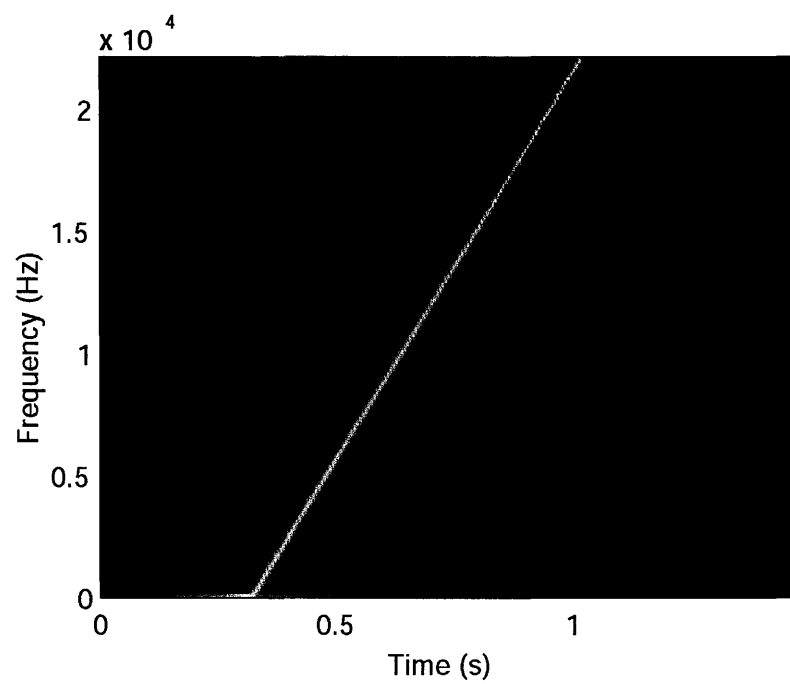


図 5.17 測定に用いた Warped-TSP のスペクトログラム (測定環境 2)

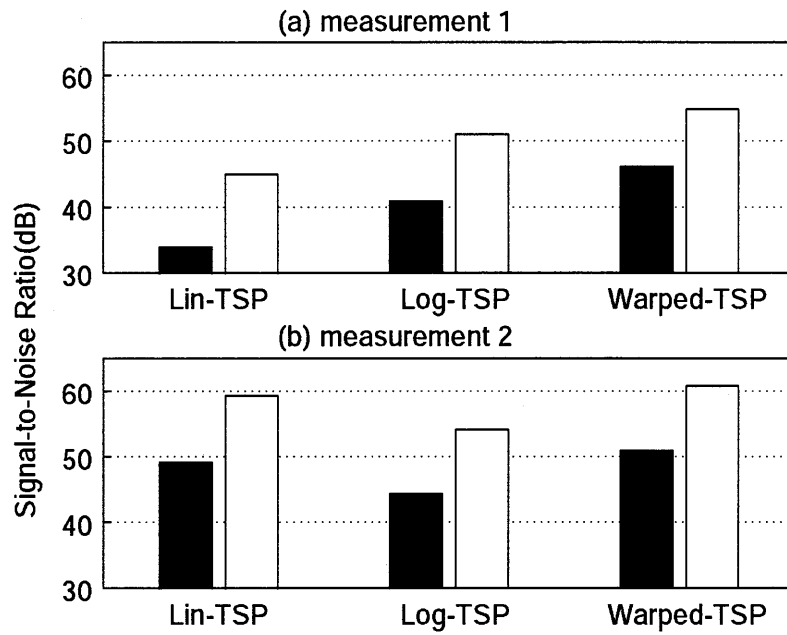


図 5.18 測定環境 1(a), および測定環境 2(b)における, 同期加算した場合 (白) としない場合 (黒) の各測定用信号による SNR.

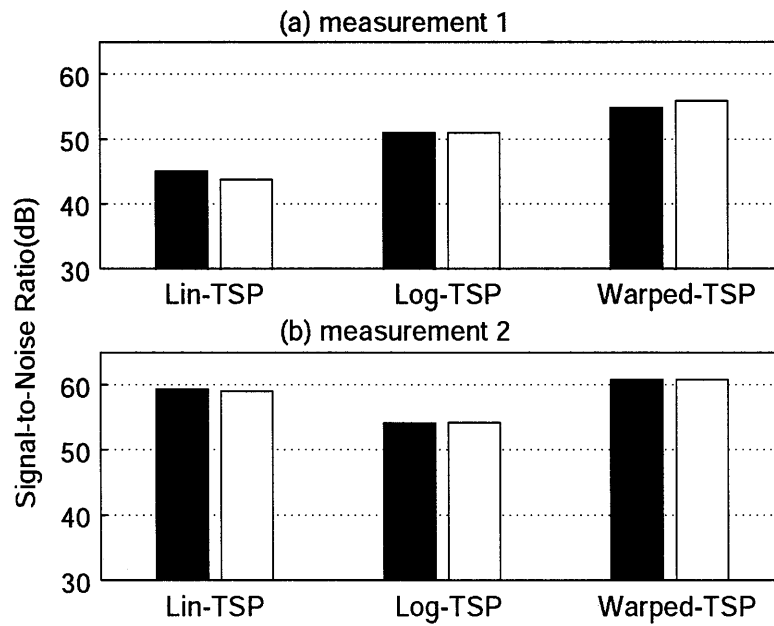


図 5.19 環状畳み込み (黒) と直線状畳み込み (白) を用いた場合の SNR.

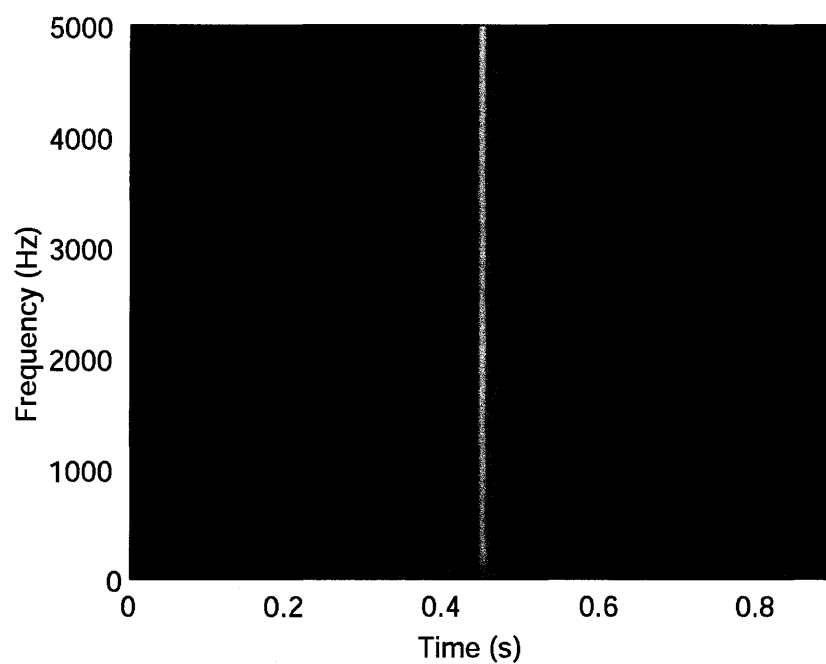


図 5.20 Lin-TSP を用いた測定で得られたインパルス応答と高調波歪のスペクトログラム.

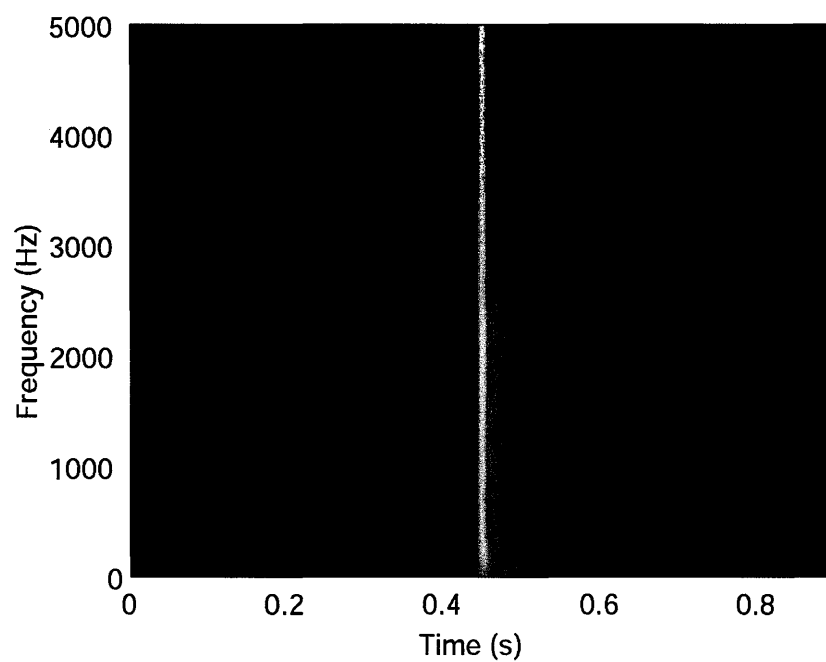


図 5.21 Log-TSP を用いた測定で得られたインパルス応答と高調波歪のスペクトログラム.

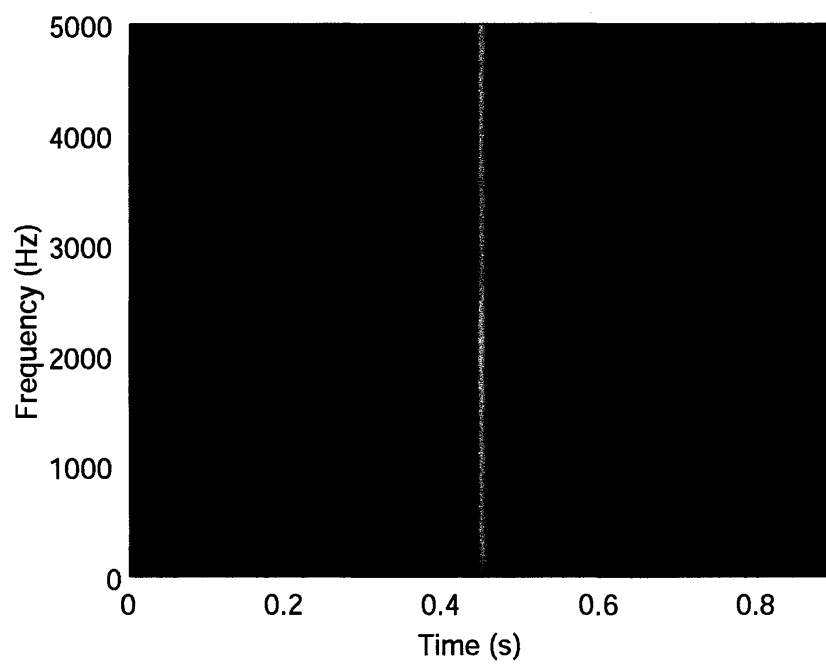


図 5.22 Warped-TSP を用いた測定で得られたインパルス応答と高調波歪のスペクトログラム.

第 6 章 結論

本論文では、音声のスペクトルからスペクトル包絡を分析時刻に依存せず推定する方法、口元から任意の観測点までの音声の伝達関数を推定する方法、音響システムのインパルス応答を環境に適応して測定できる測定用信号を提案し、検討を重ねた。

第2章では、高品質な音声分析合成システムの構築を目的として、音声スペクトルから分析時刻に依存せずスペクトル包絡を推定できる方法TANDEMの提案を行った。従来法であるSTRAIGHTの相補的時間窓と計算機シミュレーションにより比較を行い、TANDEMのほうが優れていることを示した。

まず、音声を周期信号をみなした場合のスペクトルについて論じた。音声は、声帯の振動間隔の情報と、声帯振動、口の形、舌の位置に依存して変化するスペクトル包絡で近似できることを示した。声帯振動をパルスとし、声帯振動の時間間隔を等間隔に固定したパルス列とみなした場合、時間軸上の音声は、基本周期で繰り返されるパルス列と、スペクトル包絡を逆フーリエ変換することで得られるインパルス応答との畳み込みにより表現できることを示した。インパルス応答とパルス列との畳み込みは、インパルス応答のスペクトルを離散化することを示した。これらの問題設定において、本論文で解くべき課題は、離散化されたインパルス応答のパワースペクトルを分析時刻に依存しない形の近似であることを示した。

次に、周期信号を分析する場合の問題点について論じた。はじめに、計算機で信号を分析する場合に用いられる短時間フーリエ変換について説明した。有限の長さの窓関数で周期信号を切り出すことが、時間周波数分析に与える影響を示した。具体的には、周期信号を任意の時刻にシフトした窓関数で切り出す場合、隣り合う調波の干渉が原因となり、切り出す時刻に応じて得られるパワースペクトルが異なることを示した。この問題は、窓関数で切り出すことが原因で起こる本質的な問題のため、従来行われている単一の窓関数による分析では解決できない。また、周期信号を短時間フーリエ変換を用いて時間周波数分析した場合、これらの影響が、時間・周波数両方向に周期的な谷として現れることを示した。

この問題を解決する方法として、分析時刻による影響を受けにくく、周期信号の基本

周波数の整数倍において、スペクトル包絡本来の値を保持できる分析法 TANDEM の提案を行った。TANDEM では、時間的に異なる位置に配置した 2 つの窓関数のパワースペクトルを平均することで、分析時刻に依存した項を相殺する分析法である。まず、任意の時刻にシフトした単一の窓関数で周期信号を切り出した場合に生じる干渉項について説明した。TANDEM では、単一の窓関数で生じる分析時刻へ依存する干渉項が、基本周期の半分の時刻だけ分析時刻の異なる 2 つ目の窓関数のパワースペクトルを加算することにより相殺出来ることを証明した。TANDEM を用いることで、時間・周波数両方向に生じていた周期的な変動成分から時間方向の変動が取り除けることを示した。さらに、周波数方向の変動を取り除く後処理についても説明を行った。この 2 つの手順により、現状の STRAIGHT とほぼ等価な時間周波数分析が可能になることを示した。

次に TANDEM の性質について説明を行った。まず、TANDEM 窓には任意の窓関数を用いることが出来るが、TANDEM 窓に適した窓関数の条件としてサイドローブが小さい必要があることを示した。また、音声分析では高い時間分解能の分析が要求されるため、サイドローブの小ささと高い時間分解能とを併せ持つ TANDEM 窓を選定する必要性を論じた。これらの要求を満たす TANDEM 窓を明らかにすることを狙い、計算機シミュレーションにより時間分解能・時間変動の関係を調査した。

はじめに、単一の窓を用いたピッチ同期分析による結果、STRAIGHT における相補的時間窓による結果と TANDEM による結果を比較し、適切な長さの Blackman 窓、Nuttall 窓を用いることで、STRAIGHT における相補的時間窓より、時間分解能・時間変動両方で優れた TANDEM 窓が存在することを示した。次に、周期信号に雑音が含まれている場合について同様の実験を行い、TANDEM 窓は耐雑音性に関しても相補的時間窓より優れていることを示した。さらに、TANDEM が必要とする基本周期に誤差が含まれている場合についても検討し、TANDEM は基本周期の誤差に対しても頑健であることを明らかにした。これらの結果に基づいて TANDEM に適した窓関数が Blackman 窓、Nuttall 窓であることを示した。

第 3 章では、音声の口元から観測点における伝達関数を推定するための枠組みを提案し評価を行った。

はじめに、クロススペクトル法を用いた伝達関数の推定原理について説明した。このクロススペクトル法を応用することで、音声の口元から観測点までの伝達関数を推定できることを示した。実環境で行われた収録実験に基づいて伝達関数推定の手順を説明した。まず、音声の収録方法について説明し、音声を用いた場合に得られる伝達関数は、低域と高域について SNR が低く信頼性が低いことを示した。低域の信頼性向上に関しては、全

ての信号が周波数 0 Hz で位相差が 0 になるという拘束条件を用いた後処理を示した。この後処理により、低域の雑音が大幅に抑圧できることを示した。

口元から耳元までの伝達関数を同一条件で 186 回測定し、周波数毎の標準偏差を調べることで、各周波数における信頼度を示した。その結果、10 kHz を超える高域では標準偏差が大きく、信頼度が低下していることを明らかにした。

この問題を解決するため、インパルス応答から時間周波数領域で必要な成分のみ取り出す方法の検討を行った。まず、口元から放射される音声は体を回折して伝わることで、回折は低い周波数で起こりやすく、高域において起きにくいことから、伝達特性を逆フーリエ変換して得られるインパルス応答の時間周波数表現において、取り出すべきインパルス応答の直接音は、持続時間が低域では長く、高域では短いことに着目した。周波数と持続時間が反比例の関係にあると仮定し、時間周波数領域から直接音に起因する時間周波数領域のみを選択する方法の説明を行った。具体的には、インパルス応答の時間軸を対数となるように伸縮を行い、低域通過フィルタを通して元の時間軸へ戻す方法を示した。この処理を行うことで、反射や残響、不要な暗騒音成分を除去できることを示した。無響室で行われた伝達特性推定実験結果に提案法を適用することで、パワースペクトルで重み付けした群遅延に関して標準偏差が処理前の 33 % 程度まで減少できることを示した。一方振幅周波数特性に関しても群遅延よりは効果が少ないが標準偏差が減少することを明らかにした。

第 4 章では、第 3 章で示したクロススペクトル法に基づくインパルス応答推定に適した窓関数の選定について検討を行った。

従来行われてきた研究では、クロススペクトル法に用いる信号はホワイトノイズであったが、音声の伝達関数を推定するためには、音声そのものを用いる必要がある。ホワイトノイズを測定用信号とした場合に推定誤差がもっとも少ない窓関数は提案されているが、音声のように周期的でスペクトルも平坦ではない信号を測定用信号とした場合についての検討はされていなかった。

そこで、これまで検討されてきた計算機シミュレーションを、測定用信号が音声の場合について同様に行うことで、音声を測定用信号とした場合に推定誤差を低減できる窓関数の条件を明らかにした。ホワイトノイズにおける結果と対比するため、ホワイトノイズ以外にも、音声、ピンクノイズ、短時間スペクトルが音声で周期性が存在しない雑音について比較を行い、測定用信号の狭い帯域におけるパワースペクトルの変動量が推定誤差に影響していることを計算機シミュレーションにより示した。このような信号を用いて伝達関数の推定を行った場合、サイドローブの小さい Hanning 窓や Blackman 窓

の推定誤差が小さいことを明らかにした。

さらに、狭い帯域におけるパワーの変動量を 0 dB から 60 dB まで変更して同様のシミュレーションを行った。サイドローブの小さい窓関数の推定誤差は、変動量が大きくなるに従い、他の窓関数と比較して相対的に小さくなることを示した。これらの計算機シミュレーション結果より、一般的な音声で生じる程度のパワー変動ではホワイトノイズの場合に最適とされた窓関数ではなく、Hanning 窓や Blackman 窓のようにサイドローブの小さい窓関数が適していることを明らかにした。

第 5 章では、音響機器から放射される音のインパルス応答測定を目的とし、測定環境に依存せず高い精度でインパルス応答が測定可能な測定用信号について検討を行った。

まず、従来行われているインパルス応答測定法について説明を行い、測定用信号を用いたインパルス応答が今回の目的に適していることを示した。これまでに測定用信号として提案されている信号を紹介し、それぞれの長所・短所について説明した。また、現存する測定用信号では、測定環境に応じて得られる精度の優劣が異なることを示した。高い精度で測定を行うためには、測定環境の特性を予め測定し、その上で測定環境に適した信号を選択しなければならないことを示した。

これらの要望を満たすため、測定環境の特性を入力することで、その環境に適した測定用信号を設計する方法の提案を行った。測定環境として多数の環境で暗騒音の特性測定を行い、一般的な測定環境は、提案した測定用信号に適していることを示した。また、提案法は、入力信号から観測点において収録された過程全てに含まれる高調波歪みの影響が受け難いことを証明した。

暗騒音の特性が異なる 2 箇所でインパルス応答の測定用実験を行った。提案手法は暗騒音に基づいて適切に設定を行い、測定精度を指標として従来用いられてきた測定用信号と比較した。その結果、提案手法は測定環境によらず、従来法より高い精度でインパルス応答を測定できることが明らかとなった。また、高調波歪みに関しても、従来法より影響されにくいことを示した。

以上のことから、提案法を応用することで高品質の音声分析合成システムを構成することが可能であると考えられる。また、口元から任意の場所への伝達関数推定、任意の音響機器の特性計測が可能となった。

今後の課題として、まず、音声から基本周波数情報をより精度良く推定する方法、非周期性成分の高精度推定法の検討があげられる。これは、本論文で提案した TANDEM に基づく方法を検討している。また、相補的時間窓の代わりに TANDEM を用いることを提案しているが、合成音声の主観評価が行われていないため、実験を行う必要がある。

伝達関数推定では、ダミーヘッドの放射パターンと人間の放射パターンとの比較を行う必要がある。この比較結果に基づき、ダミーヘッドを用いて行われてきた様々な評価の妥当性について検証を行う予定である。

参考文献

- 1) A.V. Oppenheim, R.W. Schafer 著, 伊達玄 訳, “デジタル信号処理 (上), ” コロナ社, 1978.
- 2) Michael Unser and Akram Aldroubi, “A general sampling theory for nonideal acquisition devices,” IEEE Trans. Signal Processing, vol.42, no.11, pp.2915-2925, 1994.
- 3) Michael Unser, “Sampling – 50 years after Shannon,” Proceedings of the IEEE, vol.88, no.4, pp.569-587, 2000.
- 4) 辻井重男 “デジタル信号処理の基礎, ” コロナ社, 1988.
- 5) W.T. Cochran, J.W. Cooley, D.L. Favin, H.D. Helms, R.A. Kaenel, W.W. Lang, G.C. Maling, D.E. Nelson, C.M. Rader, P.D. Welch, “What is the Fast Fourier Transform?” IEEE Trans. Audio and Electroacoustics, vol.AU-15, no.2, pp.45-55, June 1967.
- 6) J.W. Cooley, P.A.W. Lewis, P.D. Welch, “The Finite Fourier Transform, ” IEEE Trans. Audio and Electroacoustics, vol.AU-17, no.2, pp.77-85, June 1969.
- 7) P.D. Welch, “The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, Modified periodograms,” IEEE Trans. on audio and electroacoustics, vol.AU-15, no.2, pp.70-73, June, 1967.
- 8) S.M. Kay, S.L. Marple, JR., “Spectrum analysis— A modern perspective,” Proceedings of the IEEE, vol.69, no.11, pp.1380-1419, Nov. 1981.
- 9) 金井浩 “音・振動のスペクトル解析, ” コロナ社, 1999.
- 10) 東山三樹夫, 白井克彦, “信号解析とデジタル処理, ” 培風館, 1999.
- 11) C. Bingham, M.D. godfrey, J.W. Tukey, “Modern techniques of power spectrum estimation,” IEEE Trans. Audio and electroacoustics, vol.AU-15, no.2, June 1967.
- 12) Jae S. Lim, Alan V. Oppenheim, 青山友紀訳 “現代デジタル信号処理理論とその応用, ” 丸善株式会社, 1992.
- 13) A.V. Oppenheim, R.W. Schafer, “Discrete-time signal processing,” Prentice Hall, 1999.
- 14) L. Cohen, “Time-frequency analysis,” Prentice Hall, Englewood Cliffs, NJ, 1995.
- 15) C.G. Carter, C.H. Knapp, A.H. Nuttall, “Estimation of the magnitude-squared coherence function via overlapped fast Fourier transform processing,” IEEE Trans. of Audio and Electroacoustics, vol.AU-21, no.4, pp.337-344, 1983.

- 16) Fredric J. Harris, "On the use of windows for harmonic analysis with the discrete Fourier transform," *Proceedings of the IEEE*, vol.66, no.1, pp.51-83, Jan. 1978.
- 17) Albert H. Nuttall, "Some windows with very good sidelobe behavior," *IEEE Transactions on acoustics, speech, and signal processing*, vol.ASSP-29, no.1, Feb. 1981.
- 18) 今井聖, "音声信号処理," 森北出版株式会社, 1996.
- 19) H. Dudley, "Remaking speech," *J. Acoust. Soc. Am.*, vol.11, no.2, pp.169-177, 1939.
- 20) M.R. Schroeder, "Vocoders: Analysis and synthesis of speech," *Proc. IEEE*, vol.54, no.5, pp.720-734, 1966.
- 21) L.R. Rabiner, M. Cheng, A. Rosenberg, C. McGonegal, "A comparative performance study of several pitch detection algorithms," *IEEE Transactions on signal processing*, vol.24, no.5, pp-399-418, 1976.
- 22) T. Shimamura, H. Kobayashi, "Weighted autocorrelation for pitch extraction of noisy speech," *IEEE Transactions on speech and audio processing*, vol.9, no.7, pp.727-730, 2001.
- 23) M.J. Ross, H.L. Shaffer, A. Cohen, R. Freudberg, H.J. Manley, "Average magnitude difference function pitch extractor," *IEEE Transactions on acoustics, speech, and signal processing*, vo. ASSP-22, no.5, 1974.
- 24) A. Cheveigne and H. Kawahara, "YIN, a dundamental frequency estimator for speech and music," *J. Acoust. Soc. Am.*, vol.111, pp.1917-1930, 2002.
- 25) A.M. Noll, "Short-time spectrum and "cepstrum" techniques for vocal pitch detection," *J. Acoust. Soc. Am.*, vol.36, no.2, pp.296-302, 1964.
- 26) A.V. Oppenheim and Schafer. R.W, "Homomorphic analysis of speech," *IEEE Trans. Audio, Electroacoust.*, vol.AU-16, no.2, pp.221-226, 1968.
- 27) A.M. Noll, "Cepstrum pitch determination," *J. Acoust. Soc. Am.*, vol.41, no-2, pp.293-309, 1967.
- 28) 阿竹義徳, 入野俊夫, 河原英紀, 陸金林, 中村哲, 鹿野清宏, "調波成分の瞬時周波数を用いた基本周波数推定方法," *信学論*, vol.J83-D-II, pp.2077-2086, 2000.
- 29) 河原英紀, "Vocoder のもう一つの可能性を探る," *日本音響学会誌*, vol.63, no.8, pp.442-449, 2007.
- 30) H. Kawahara, Alain de Cheveigne, H. Banno, T. Takahashi and T. Irino, "Nearly defect-free F0 trajectory extraction for expressive speech modifications based on STRAIGHT," *Proc. Interspeech2005*, pp.537-540, Sept. 2005.

- 31) T. Nakatani and T. Irino, "Robust and accurate fundamental frequency estimation based on dominant harmonic components," J. Acoust. Soc. Am., vol.116, no.6, pp.3690-3700, Dec. 2004.
- 32) 板倉文忠, 斉藤収三, "統計的手法による音声スペクトル密度とホルマント周波数の推定," 電子情報通信学会論文誌, vol.53-A, no.1, pp.35-42, 1970.
- 33) 板倉文忠, 斉藤収三, "最尤スペクトル推定法を用いた音声情報圧縮," 日本音響学会誌, vol.27, no.9, pp.17-26, 1971.
- 34) B.S. Atal and S.L. Hanauer, "Speech analysis and synthesis by linear prediction of the speech wave," J. Acoust. Soc. Am., vol.50, no.2, pp.637-655, 1971.
- 35) L.R. Rabiner and R.W. Schafer, "Digital processing of speech signals," Prentice Hall, Inc.,
- 36) A.V. Oppenheim, "Speech analysis-synthesis system based on homomorphic filtering," J. Acoust. Soc. Am., vol.45, no.2, pp.458-465, 1969.
- 37) 今井聖, 北村正, "対数振幅特性近似フィルタを用いた音声の分析合成系," 信学論A, vol.J61-A, no.6, pp.527-534, 1978.
- 38) B. Gold and C.M. Rader, "The channel vocoder," IEEE Trans. Audio, Electroacoust., vol.AU-15, no.4, pp.148-161, 1967.
- 39) 板倉文忠, "スペクトル符号化に基づく音声分析合成," 日本音響学会誌, vol.37, no.5, pp.197-203, 1981.
- 40) H. Kawahara, "Speech representation and transformation using adaptive interpolation of weighted spectrum: VOCODER revisited," ICASSP'97, vol.2, pp.1303-1306, 1997.
- 41) 河原英紀, "聴覚の情景分析が生んだ高品質 VOCODER: STRAIGHT," 日本音響学会誌, vol.54, pp.521-526, 1998.
- 42) H. Kawahara, I. Masuda-Katsuse, and A. Cheveigne, "Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction," Speech Commun., vol.27, pp.187-207, 1999.
- 43) 河原英紀, P. Zolfaghari, A. Cheveigne and R.D. Patterson, "周波数から瞬時周波数への写像の不動点を用いた音源情報の抽出について," 信学技報, SP-99-40, pp.1-8, 1999.
- 44) H. Kawahara, Jo Estill and O. Fujimura, "Aperiodicity extraction and control using mixed mode excitation and group delay manipulation for a high quality speech analysis, modification and synthesis system STRAIGHT," MAVEBA 2001, Sept.13-15, Firentze Italy, 2001.

- 45) 河原英紀, 森勢将雅, 高橋徹, 入野俊夫, 坂野秀樹, 藤村靖, “STRAIGHT スペクトルに基づく音源信号の抽出と非周期性成分の評価について,” 信学技報, SP-2006-83, pp.43-48, 2006.
- 46) Hideki Kawahara, Hideki Banno, Masanori Morise, Yumi Hirachi, “A cappella synthesis demonstrations using RWC music database,” Proc. NIME04, pp.130-131, 2004.
- 47) H. Kawahara and H. Matsui, “Auditory morphing based on an elastic perceptual distance metric in an interference-free time-frequency representation,” ICASSP 2003, pp.256-259, 2003.
- 48) 坂野秀樹, 武田一哉, 板倉文忠, 鹿野清宏, “包絡と音源の独立操作による音声モーフィング,” 電子情報通信学会論文誌 A, vol.J81-A, no.2, pp.261-268, 1998.
- 49) 内村佳典, 坂野秀樹, 板倉文忠, “声道断面積関数による声質制御方式に関する検討,” 日本音響学会大会講演秋, pp.419-420, 2007.
- 50) Hideki Banno, Hiroaki Hata, Masanori Morise, Toru Takahashi, Toshio Irino, Hideki Kawahara, “Implementation of realtime STRAIGHT speech manipulation system,” Acoust. Sci. & Tech. vol.28, no.3, pp.140-146, 2007.
- 51) 畑宏明, 高橋徹, 坂野秀樹, 入野俊夫, 河原英紀, “聴覚特性を考慮した STRAIGHT スペクトル補間特性とその主観評価について,” 日本音響学会 2006 年秋季研究発表会講演論文集, pp.271-272, 2006.
- 52) 畑宏明, 高橋徹, 入野俊夫, 河原英紀, “聴覚特性を考慮した高品質領域における STRAIGHT パラメタの補間特性に関する検討,” 電子情報通信学会技術研究報告, SP2006-139, Jan. 26-27, 2006.
- 53) J.L. Flanagan, “Analog measurements of sound radiation from the mouth,” J. Acoust. Soc. AM., vol.32, No.12, pp.1613-1620, Dec, 1960.
- 54) Katz Brian FG, d’Alessandro Christophe, “Directivity measurements of the singing voice,” Proc ICA2007, 2007.
- 55) Cabrera Densil, Davis Pamela J, Connolly Anna, “Vocal directivity of eight opera singers of spectro-spatial parameters,” Proc ICA2007, 2007.
- 56) 貫名真澄, 河原英紀, “発話時の頭部周辺での音声の伝達特性について,” 音響誌, vol.59, no.5, pp.256-260, May, 2003.
- 57) M.V. Mathews, Joan E. Miller, and E.E. David, “Pitch synchronous analysis of voiced sounds,” J. Acoust. Soc. Am., vol.33, no.2, pp.179-185, 1961.
- 58) 森勢将雅, 高橋徹, 河原英紀, 入野俊夫, “窓関数による分析時刻の影響を受けにくい周期信号のパワースペクトル推定法,” 電子情報通信学会論文誌 D(2007 年 12 月掲載予定).

- 59) 河原英紀, 森勢将雅, 高橋徹, 西村竜一, 入野俊夫, 坂野秀樹, “時間方向および周波数方向の周期性の影響を除去した周期信号のパワースペクトルの表現について,” 電子情報通信学会技術研究報告, SP2007-26, vol.37, no.6, pp.395-400, July, 2007.
- 60) ITU-T series p.51, Telephone transmission quality, Objective measuring apparatus, Artificial Mouth.
- 61) ITU-T series p.58, Telephone transmission quality, Objective measuring apparatus, Head and torso simulator for telephonometry.
- 62) 党建武, 中井孝芳, 鈴木久喜, “連続音声発話時の鼻腔からの放射音圧の文脈による変化の推定,” 電子情報通信学会論文誌, vol.J74-A, no.7, pp.1121-1124, July, 1991.
- 63) 鈴木久喜, 党建武, 中井孝芳, “音声発声時の口唇, 鼻孔, 喉頭部の音と振動の推定および非鼻音化音声における鼻腔口腔間の音響漏洩のシミュレーション,” 電子情報通信学会論文誌, vol.J74-A, no.12, pp.1705-1714, Dec, 1991.
- 64) T. Chiba, M. Kajiyama, “The vowel its nature and stracture,” Tokyo-Kaiseikan Pub.Co.Ltd, 1942.
- 65) Fant.G, “Acoustic theory of speech production,” Monton & Co.’s- Gravenhage, 1960.
- 66) M. Morise, T. Irino, H. Kawahara, “Accuracy improvement in speech sound propagation measurement using logarithmic temporal manipulation,” Proc. Eusipco2006, Italy, Sept, 2006.
- 67) 柏木潤, M 系列とその応用, 昭晃堂, 東京, 1996.
- 68) N. Aoshima, “Computer-generated pulse signal applied for sound measurement,” J. Acoust. Soc. AM., vol.69, no.5, pp.1484-1488, May, 1981.
- 69) 永田仁史, 劉家祥, 安部正人, 城戸健一, “クロススペクトル法によるコヒーレンス関数及びインパルスレスポンス推定の際の窓関数の影響,” 音響誌, vol.45, no.12, pp.950-956, Dec, 1989.
- 70) 福島学, 城戸健一, “クロススペクトル法によるインパルスレスポンス推定における推定誤差の検討,” 音響誌, vol.55, no.4, pp.265-274, April, 1999.
- 71) K. Kido, H. Suzuki, T. Ono and M. Fukushima, “Deformation of impulse reponse estimates by time window in cross spectral technique,” J. Acoust. Soc. Jpn. (E), vol.19, no.5, pp.349-361, 1998.
- 72) 福島学, 城戸健一, “インパルスレスポンス推定に適した時間窓の検討,” 音響誌, vol.56, no.2, pp.78-84, Feb, 2000.
- 73) 城戸健一. 基礎音響工学, コロナ社, 東京, 1990.

- 74) John R.Deller, John H.L.Hansen, John G.Proakis, "Discrete-time processing of speech signals," The Institute of Electrical and Electronics Engineers Inc., 2000.
- 75) 森勢将雅, 河原英紀, "対数時間軸伸縮と多点計測を用いた音響測定法の誤差評価について," 信学技報, EA2003-53, pp.17-22, August, 2003.
- 76) H. Alrutz and M.R. Schroeder, "A fast Hadamard transform method for the evaluation of measurements using pseudorandom test signals," Proc. 11th ICA, Paris, vol.6, pp.235-238, July, 1983.
- 77) A.J. Berkhout, D.de Vries, and M.M. Boone, "A new method to acquire impulse responses in concert halls," J. Acoust. Soc. AM., vol.68, no.1, pp.179-183, Jan, 1980.
- 78) J. Schoukens, R. Pintelon, E.Van Der Ouderaa, and J. Renneboog, "Survey of Excitation Signals for FFT based Signal Analyzers," IEEE Trans. Instrumentation and Measurement, vol.37, no.3, pp.342-352, Sept, 1988.
- 79) 佐藤史明, "室内音響インパルス応答の測定技術," 音響誌, vol.58, no.10, pp.669-676, Oct, 2002.
- 80) 金田豊, "M 系列を用いたインパルス応答測定における誤差の実験的検討," 音響誌, vol.52, no.10, pp.752-759, Oct, 1996.
- 81) 藤元卓也, "低域バンドでの SN 比改善を目的とした TSP 信号に関する検討," 音講論, pp.433-434, Sept, 1999.
- 82) 伊藤次男, "音響測定方法およびその装置," 公開特許公報, 特開平 5-118906, 1993.
- 83) A. Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique," 108th AES Convention, 5093,(D-4), Feb, 2000.
- 84) 藤元卓也, "低域バンドでの SN 比改善を目的とした TSP 信号に関する検討 —高調波歪の除去—," 音講論, pp.555-556, March, 2000.
- 85) 守谷直也, 金田豊, "Logarithmic TSP 信号における非線形高調波歪に関する検討," 音講論, pp.637-638, March, 2004.
- 86) 守谷直也, 池田亜希, 金田豊, "インパルス応答計測におけるスピーカの非線形歪みに関する検討," 音講論, pp.735-736, Sept, 2004.
- 87) 守谷直也, 金田豊, "インパルス応答計測のための最適信号の検討," 信学技報, EA2004-136, pp.1-6, March, 2005.
- 88) 鈴木陽一, 浅野太, 金学胤, 曾根敏夫, "時間引き伸ばしパルスの設計法に関する考察," 信学技法, EA92-86, pp.17-24, Dec, 1992.

- 89) 金学胤, 浅野太, 鈴木陽一, 曾根敏夫, “TSP を用いて設計長より長いインパルス応答を計測する場合の計測誤差について,” 音講論, pp.451–452, March, 1993.
- 90) Yoiti Suzuki, Futoshi Asano, Hack-Yoon Kim, and Toshio Sone, “An optimum computer-generated pulse signal suitable for the measurement of very long impulse responses,” J. Acoust. Soc. AM., vol.97, no.2, pp.1119–1123, Feb, 1995.

研究業績

学術論文

- 河原 英紀, 生駒 太一, 森勢 将雅, 高橋 徹, 豊田 健一, 片寄 晴弘, “モーフィングに基づく歌唱デザインインタフェースの提案と初期検討,” 情報処理学会論文誌, vol.48, no.12, pp.3637-3648, 2007.
- 森勢将雅, 高橋徹, 河原英紀, 入野俊夫, “窓関数による分析時刻の影響を受けにくい周期信号のパワースペクトル推定法,” 電子情報通信学会 論文誌 D, vol.J90-D, no.12, pp.3265-3267 Dec. 2007.
- 森勢将雅, 田原佳代子, 高橋徹, 入野俊夫, 河原英紀, “単母音による歌唱音声スペクトルの統計的分析に基づく音色制御法の提案と評価,” FIT2007 第6回情報科学技術フォーラム, LE-004, pp.119-122, 2007.
- 森勢将雅, 入野俊夫, 河原英紀, “測定用信号として音声を用いたクロススペクトル法によるインパルス応答推定の誤差評価,” 電子情報通信学会論文誌 A vol.J90-A, no.7, pp.559-566, July, 2007.
- Hideki Banno, Hiroaki Hata, Masanori Morise, Toru Takahashi, Toshio Irino, Hideki Kawahara, “Implementation of realtime STRAIGHT speech manipulation system,” Acoust. Sci. & Tech. vol.28, no.3, pp.140-146, 2007.
- 森勢将雅, 入野俊夫, 坂野秀樹, 河原英紀, “暗騒音と高調波ひずみに頑健なインパルス応答測定用信号: Warped-TSP,” 電子情報通信学会論文誌 Vol.J89-A, No.1, pp.7-14, Jan. 2006.
Masanori Morise, Toshio Irino, Hideki Banno, Hideki Kawahara, “Warped-TSP: An acoustic measurement signal robust to background noise and harmonic distortion,” Electronics and Communications in Japan (Part III: Fundamental Electronic Science), vol.90, no.4, pp.18-26, April, 2007. (Translation paper)

国際会議

- Hideki Kawahara, Masanori Morise, Toru Takahashi, Hideki Banno, Toshio Irino, Osamu Fujimura, “Group delay for acoustic event representation and its application for speech aperiodicity analysis,” 15th European Signal Processing Conference (EUSIPCO 2007), pp.2219-2223, Poznan, Poland, 3-7, Sept. 2007.

- Masanori Morise, Toshio Irino, Hideki Banno, Hideki Kawahara, “Warped-time-stretched pulse: An acoustic test signal robust against ambient noise,” J. Acoust. Soc. Am., vol.120, no.5, pp.3223, Nov, 2006. (ASA and ASJ joint meeting, Hawaii) [Invited talk]
- Toru Takahashi, Masanori Morise, Toshio Irino, “Evaluating naturalness of speeches morphed by independently using the interpolation ratios of the time-frequency axes and amplitude,” J.Acoust.Soc.Am., vol.120, no.5, pp.3036, Nov, 2006. (ASA and ASJ joint meeting, Hawaii)
- Masanori Morise, Toshio Irino, and Hideki Kawahara, “Logarithmic temporal processing applied to accurate empirical transfer function measurements in vocal sound propagation,” Proc. EUSIPCO2006, CD-ROM, Sept. 2006.
- Masanori Morise, Toshio Irino, Hideki Banno, and Hideki Kawahara, “A test signal robust against background noise in the measurement of acoustic impulse responses: Warped-TSP,” The 34th International Congress and Exposition on Noise Control Engineering (Internoise 2005) , Rio de Janeiro, Brazil, 7-10 Aug. 2005.
- Masanori Morise and Hideki Kawahara, “Loudspeaker equalization based on multi-location observation with reliable time-frequency region selection and its evaluation using sound propagation measurement,” Proc. EUSIPCO’2004 Vienna, pp.1995-1998, 2004.
- Hideki Kawahara, Hideki Banno, Masanori Morise, Yumi Hirachi, “A cappella synthesis demonstrations using RWC music database,” Proc. NIME04, pp.130-131, 2004.
- Masanori Morise, Hideki Kawahara, “A new acoustic measurement and compensation method based on logarithmic transformation of the time axis and multi-location acquisition,” International Congress on Acoustics Kyoto, vol.1, pp.721-724, 2004.
- Masanori Morise, Hideki Kawahara, “Logarithmic temporal axis manipulation and its application for measuring perceptually salient acoustic features of loudspeakers based on multiple observations,” J. Acoust. Soc. Am., 114, pp. 2460, 2003.

査読付き国内学会

- Masanori Morise, Shunsuke Fukuda, Toru Takahashi, Toshio Irino, Hideki Kawahara, “A unified designed criteria for noise adaptive sound reproduction system based on an auditory model,” AES 13th Regional Convention, CD-ROM, Tokyo 2007.

- 河原英紀, 生駒太一, 森勢将雅, 高橋徹, 豊田健一, 片寄晴弘, “歌唱モーフィングに基づく音質と歌い回し転写の知覚的検討,” インタラクション 2007, 一般講演論文, CD-ROM, 2007.

国内学会

- 河原英紀, 森勢将雅, 高橋徹, 西村竜一, 坂野秀樹, 入野俊夫, “TANDEM-STRAIGHTに基づく基本周波数の抽出と評価について,” 情報処理学会研究報告, vol.2007, no.129, pp.259-264, Dec, 2007.
- 河原英紀, 森勢将雅, 高橋徹, 西村竜一, 坂野秀樹, 入野俊夫, “分析位置に依存しない周期信号のパワースペクトル推定法に基づく音声分析変換合成法 STRAIGHTの再構成について,” 電子情報通信学会, 第22回信号処理シンポジウム, pp.310-315, November, 2007. (東北大学/宮城).
- 森勢将雅, 高橋徹, 河原英紀, 入野俊夫, “周期信号の分析時刻に依存しないパワースペクトル推定法における対雑音性の評価,” 電子情報通信学会, 第22回信号処理シンポジウム, pp.581-586, November, 2007. (東北大学/宮城).
- 高橋 徹, 大西 壮登, 森勢 将雅, 河原 英紀, 入野 俊夫, “双方向変換により共通化された時間周波数軸上でのパラメタ混合に基づく音声モーフィング,” 電子情報通信学会, 第22回信号処理シンポジウム, pp.316-321, November, 2007. (東北大学/宮城).
- 吉田有里, 森勢将雅, 高橋徹, 河原英紀, “ポップス系歌唱音声の STRAIGHT により分析とスペクトル変動の統計的性質について,” 信学技報, vol.107, no.282, pp.31-36, Oct. 2007.jbrj
- 河原英紀, 森勢将雅, 高橋徹, 西村竜一, 入野俊夫, 坂野秀樹, “時間方向および周波数方向の周期性の影響を除去した周期信号のパワースペクトルの表現について,” 信学技報, SP2007-26, vol.37, no.6, pp.395-400, July 26-27, 2007.
- 高橋徹, 森勢将雅, 大西壮登, 西村竜一, 入野俊夫, 坂野秀樹, 河原英紀, “音声テクスチャーマッピングに基づく変換関数を利用した音声モーフィングの自動化について,” 信学技報, SP2007-6, vol.107, no.77, pp.31-34, May 31, 2007.
- 高橋徹, 森勢将雅, 西村竜一, 入野俊夫, 河原英紀, 坂野秀樹, “母音情報に基づく話者変換システムの提案,” 電子情報通信学会技術研究報告, SP2006-162, vol.106, no.613, pp.13-18, March 26-27, 2007.
- 福田俊介, 森勢将雅, 河原英紀, 入野俊夫, “騒音適応型音楽再生システムのパラメタ設定における聴覚モデルによる評価の応用について,” 電子情報通信学会・日本音響学会, EA2006-77, pp.43-47, Nov, 2006.

- 森勢将雅, 高橋徹, 入野俊夫, 河原英紀, “知覚特性を考慮したインパルス応答補償法についての一検討,” 電子情報通信学会・日本音響学会, EA2006-72, pp.13-18, Nov, 2006.
- 森勢将雅, 高橋徹, 入野俊夫, 河原英紀, “帯域毎の持続時間に基づく音響イベント検出と位置推定への応用,” 電子情報通信学会・日本音響学会, EA2006-73, pp.19-24, Nov, 2006.
- 河原英紀, 森勢将雅, 高橋徹, 入野俊夫, 藤村靖, “STRAIGHT スペクトルに基づく音源信号の抽出と非周期成分の評価について,” 電子情報通信学会・日本音響学会 SP2006-53, pp.43-48, Nov, 2006.
- 高橋徹, 森勢将雅, 入野俊夫, 河原英紀, “高品質音声分析変換合成システム STRAIGHT における分析窓の検討,” 電子情報通信学会・日本音響学会, SP2006-42, pp.1-6, Aug, 2006.
- 森勢将雅, 入野俊夫, 河原英紀, “対数時間軸伸縮による人間の頭部伝達関数測定の精度改善について,” 電子情報通信学会 応用音響研究会, EA2005-64, pp.43-48, Oct, 2005.
- 田原佳代子, 高橋徹, 森勢将雅, 坂野秀樹, 河原英紀, “歌唱音声制御に伴うスペクトル変動の主成分分析と合成への応用について”, Vol.105, No.198, SP2005-38, pp.19-24, July, 2005.
- 森勢将雅, 入野俊夫, 坂野秀樹, 河原英紀, “暗騒音に頑健なインパルス応答測定用信号の設計手法,” 電子情報通信学会・日本音響学会, EA2004-44, pp.37-42, Aug, 2004.
- 森勢将雅, 河原英紀, “対数時間軸伸縮と多点計測に基づく補償特性について,” 日本音響学会, 聴覚研究会, vol.33, no.8, pp.579-584, Oct, 2003.
- 森勢将雅, 河原英紀, “対数時間軸伸縮と多点計測を用いた音響測定法の誤差評価について,” 電子情報通信学会・日本音響学会 EA2003-53, pp.17-22, Aug, 2003.

大会講演

- 河原英紀, 森勢将雅, 高橋徹, 西村竜一, 入野俊夫, 坂野秀樹, “STRAIGHT における時間周波数分析の新しい定式化と実装について,” 日本音響学会 2007 年秋季研究発表会講演論文集, pp.347-348, 山梨大学, Sept 19-21, 2007.
- 森勢将雅, 高橋徹, 入野俊夫, 河原英紀, “周期信号の分析時刻に依存しないパワースペクトル推定法に適した窓関数の検討”, 日本音響学会 2007 年秋季研究発表会講演論文集, pp.349-350, 山梨大学, Sept 19-21, 2007.

- 森勢将雅, 高橋徹, 入野俊夫, 河原英紀, “聴覚フィルタを評価に用いた逆フィルタ設計法に関する一考察”, 日本音響学会 2007 年秋季研究発表会講演論文集, pp.737-738, 山梨大学, Sept 19-21, 2007.
- 高橋徹, 大西壮登, 森勢将雅, 坂野秀樹, 河原英紀, 入野俊夫, “音声モーフィングのための母音スペクトル間区分線形写像関数自動設計手法,” FIT2007 第 6 回情報科学技術フォーラム, pp.233-236, 中京大学, 愛知, 2007 年 9 月 5 日-7 日.
- 森勢将雅, 高橋徹, 入野俊夫, 河原英紀, “パルス列を用いた高域における群遅延操作の弁別閾推定,” 日本音響学会, pp.517-518, March. 2007.
- 森勢将雅, 田原佳代子, 高橋徹, 入野俊夫, 河原英紀, “STRAIGHT を用いた歌唱合成における母音スペクトル形状制御の効果について,” 日本音響学会, pp.219-220, March. 2007.
- 鈴木健太郎, 森勢将雅, 高橋徹, 河原英紀, 入野俊夫, “低周波数領域での区分線形補間の弊害についての一検討,” 日本音響学会, pp.275-276, March. 2007.
- 河原英紀, 森勢将雅, 高橋徹, 入野俊夫, “音声モーフィングにおける周波数座標変換関数の設計と知覚への影響について,” 日本音響学会, pp.477-478, March. 2007.
- 森勢将雅, 入野俊夫, 河原英紀, “対数時間軸伸縮を用いたインパルス応答測定における直接音・反射音成分の分離について,” 日本音響学会, pp.429-430, Sept. 2006.
- 高橋徹, 森勢将雅, 入野俊夫, 河原英紀, “ブラックマン窓を用いた STRAIGHT スペクトル分析,” 日本音響学会, pp.215-216, March. 2006.
- 森勢将雅, 入野俊夫, 河原英紀, “TSP を用いた音響測定における PC 用 AD/DA 変換システムの選定について,” 日本音響学会, pp.653-654, March. 2006.
- 河原英紀, 西雅史, 森勢将雅, 野口美咲, 高橋徹, 入野俊夫, “STRAIGHT に基づくモーフィングのオブジェクト化による拡張と部分モーフィングの応用について,” 日本音響学会, pp.505-506, March. 2006.
- 森勢将雅, 入野俊夫, 河原英紀, “PC 用 AD/DA 変換器における折り返し歪について,” 日本音響学会 2005 年秋季研究発表会講演論文集, pp.679-680, Sept, 2005.
- 森勢将雅, 平地由美, 坂野秀樹, 入野俊夫, 河原英紀, “STRAIGHT を用いたビブラート歌唱音声の統計的性質,” 日本音響学会 2005 年春季研究発表会講演論文集, vol.1, pp.269-270, March, 2005.
- 田原佳代子, 森勢将雅, 坂野秀樹, 入野俊夫, 河原英紀, “歌唱音声の音量変化に伴うスペクトル変形の分析について,” 日本音響学会, vol.1, pp.271-272, March. 2005.

- 森勢将雅, 入野俊夫, 坂野秀樹, 河原英紀, “暗騒音に基づいたインパルス応答測定用信号の設計手法,” 日本音響学会 2005 年秋季研究発表会講演論文集, vol.1, pp.593-594, Sept, 2004.
- 森勢将雅, 河原英紀, “非線形時間軸伸縮に基づく特性測定法の波面計測による検討,” 日本音響学会 2005 年春季研究発表会講演論文集, vol.1, pp.633-634, March, 2004.
- 森勢将雅, 河原英紀, “対数時間軸伸縮と多点計測を用いた音響測定法の誤差要因について,” 日本音響学会 2005 年春季研究発表会講演論文集, vol.1, pp.617-618, Sept, 2003.

特許

- 河原英紀, 森勢将雅, 高橋徹, 入野俊夫 “周期信号処理方法、周期信号変換方法および周期信号処理装置,” 特願 2007-187697, 出願日 2007 年 7 月 18 日 (和歌山大学).
- Masanori Morise, Toshio Irino, Hideki Banno and Hideki Kawahara, “Impulse Response Measurement Method and Device,” European Patent Application, 公開番号 1772713, (出願番号 No. 05765660.5 based on PCT/JP2006/012812), 出願日 2007 年 4 月 11 日. (和歌山大学).
- 森勢将雅, 入野俊夫, 河原英紀, 坂野秀樹, “インパルス応答測定法及び装置,” 日本国特許出願中, 特願 2004-221942, 出願日 2004 年 7 月 29 日. (和歌山大学)
- 森勢将雅, 入野俊夫, 河原英紀, 坂野秀樹, “インパルス応答測定法及び装置,” PCT 特許出願中, 国際出願番号 PCT/JP2005/012812. (和歌山大学)

表彰

- 第 22 回電子通信普及財団賞「テレコムシステム技術学生賞」. 2007 年 3 月.
- 和歌山大学大学院システム工学研究科長表彰. 2007 年 3 月.

付録

A. 基本周波数の推定に関する先行研究

ここでは、基本周波数推定法に関する先行研究を示す。

時間波形に基づいて推定する方法

もっとも簡単な基本周波数推定法として、信号の自己相関に基づいた方法が提案されている。任意のインパルス応答を、基本周期が T_0 であるパルス列に畳み込んだ周期信号の自己相関を求めると、基本周期 T_0 の整数倍において相関が高い。そのため、自己相関の原点以外で生ずる 1 番原点に近いピークの時刻が基本周期 T_0 となる。

ただし、声帯振動の周期は、完全に等間隔とはならない。そのため、音声の基本周波数推定では、音声信号を比較的短い (20 ms から 40 ms) 区間で窓関数を用いて切り出した波形から求めた自己相関を用いる。自己相関に基づく基本周波数推定では、真値の倍や半分の周波数を真値と誤推定する問題がある。このような問題の対応、あるいは雑音に対する頑健性、精度の向上を目的とし、相関から得られるパラメタに基づいて基本周期を推定する方法は、この他にも多数提案されている^{21, 22, 23, 24)}。

スペクトルに基づいて推定する方法

周期信号 $y(t)$ のスペクトル $Y(\omega)$ は、基本周波数の整数倍において振幅、位相が異なるパルス列となる。スペクトルの特徴に着目した基本周波数を推定する方法として、ケプストラム^{25, 26)} 用いた方法が提案されている。

ケプストラム $c(\tau)$ は、音声スペクトルの振幅の対数を逆フーリエ変換することで得られるパラメタである。

$$c(\tau) = \mathcal{F}^{-1}(\log |Y(\omega)|) \quad (\text{A.1})$$

\mathcal{F}^{-1} は逆フーリエ変換を表す記号である。時間軸上における畳み込みは、周波数軸上では積となる。そのため、振幅スペクトルの対数は、スペクトル包絡の対数とパルス列の足し算となる。パルス列のスペクトルは、時間波形を窓関数により切り出すことにより、隣接する周波数に漏れの影響が含まれる。以下では、パルス列のスペクトル $X(\omega)$ は周波数漏れの影響によりゼロが存在しないものとする。窓関数による切り出しの影響は 2 章で詳しく論ずる。

$$\log |Y(\omega)| = \log |P(\omega)| + \log |X(\omega)| \quad (\text{A.2})$$

ケプストラム $c(\tau)$ は、スペクトル包絡のケプストラムと周波数軸で広がりを持つパルス列のケプストラムの和となる。

$$c(\tau) = \mathcal{F}^{-1}(\log |P(\omega)|) + \mathcal{F}^{-1}(\log |X(\omega)|) \quad (\text{A.3})$$

ケプストラムは、周波数軸の応答を逆フーリエ変換したパラメタである。 τ はケフレンシーと呼ばれ、時間の次元を持つ。このとき、周期的なパルス列に起因する成分のケプストラムは、高次のケフレンシーでピークを持つ。高次のケフレンシーに生ずるピークの逆数が基本周波数となる²⁷⁾。

この他に、瞬時周波数に基づく基本周波数推定法²⁸⁾も提案されている。これらの方法を組み合わせて測定精度を向上させる方法も提案されている^{29, 30)}。特に、文献³⁰⁾で提案されている基本周波数推定法は、現存する基本周波数推定法の中で最も高い精度で基本周波数を推定できるとされている。耐雑音性の観点で優れた推定法³¹⁾も提案されており、SNR 0 dBにおいても基本周波数の抽出が可能となっている。このように、基本周波数に関して多数の推定法が提案されている。

B. 相補的時間窓

窓関数に基本周期の二倍の周期の正弦波を乗じると、窓関数の周波数領域での表現は、基本周波数の半分だけ上と下にシフトしたものの和となる。上下の成分間には π の位相差が加わるため、調波の間では、上下に隣接する調波成分が、 π だけ違った位相で混合される。そのため、パワースペクトルに含まれる時間方向の正弦波状の変動は、元のパワースペクトルの変動の逆位相となる。

相補的時間窓 $w_C(t)$ は、この性質を利用して窓関数 $w(t)$ から次式により作成される。なお、窓関数 $w(t)$ は、基本周期の長さの矩形窓との畳込みによりピッチ同期化されているものとする。基本周期を T_0 とする。

$$w_C(t) = w(t) \sin \frac{\pi t}{T_0} \quad (\text{B.1})$$

この窓を用いて求められたパワースペクトルを $P_C(\omega)$ とし、元の窓を用いて求められたパワースペクトルを $P(\omega)$ とする。次式のように、適切な係数 ξ の重みを用いてこれらを加算することで変動成分が相殺され、時間方向に変動しないパワースペクトル $P_R(\omega)$ が得られる。

$$P_R(\omega, t) = P(\omega, t) + \xi P_C(\omega, t) \quad (\text{B.2})$$

係数 ξ は、次式で表される時間方向の変動の相対的なエネルギーが最小となるように、数値的探索により求められる。

$$\rho^2(\xi) = \frac{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |B_R(\omega, t) - \overline{B_R(\omega)}|^2 dt d\omega}{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P_R(\omega, t) dt d\omega} \quad (\text{B.3})$$

ここで $B_R(\omega, t) = \sqrt{P_R(\omega, t)}$ は絶対値スペクトルを表し、 $\overline{B_R(\omega)}$ はその時間平均を表す。

現在の STRAIGHT では、矩形窓と畳み込まれる窓には、基本周波数と基本周期に関して等方的な広がり gaussian 窓を若干 (1.1 倍) 時間方向に引き延ばしたものを用いている。このように、現在の実装には、多数の調整すべきパラメタや経験的に決められたアドホックな工夫などが含まれている。そのため、STRAIGHT のスペクトル分析は、理論的な見通しの悪いものとなってしまっている。また、元の窓を正弦波により変調する操作には、周波数分解能を低下させるという副作用もある。

C. 本論文で用いた公式の導出

本論文で用いたいくつかの公式について、この章にて補足説明を行う。

パルス列のスペクトルの導出 (2 章)

まず、2 章で示した、以下の式の導出を行う。

$$x(t) = \sum_{n=-\infty}^{\infty} \delta(t - nT_0) \quad (\text{C.1})$$

$$X(\omega) = \sum_{n=-\infty}^{\infty} \delta(\omega - n\omega_0) \quad (\text{C.2})$$

このパルス列 $x(t)$ は、周期 T_0 を有する周期信号である。まず、このパルス列を複素フーリエ級数に展開する。なお、本論文では、デルタ関数 $\delta(t)$ は、以下の定義に従って用いることとする。

$$\delta(t) = 0 \quad \text{if } t \neq 0 \quad (\text{C.3})$$

$$\int_{-\infty}^{\infty} \delta(t) dt = 1 \quad (\text{C.4})$$

複素フーリエ級数展開では、周期 T_0 の信号 $f(t)$ と複素フーリエ係数 c_n との関係が以下のように与えられる。

$$f(t) = \sum_{n=-\infty}^{\infty} c_n e^{jn\omega_0 t} \quad (\text{C.5})$$

$$c_n = \frac{1}{T_0} \int_{-T_0/2}^{T_0/2} f(t) e^{-jn\omega_0 t} dt \quad (\text{C.6})$$

$$T_0 = \frac{2\pi}{\omega_0} \quad (\text{C.7})$$

対象となるパルス列 $x(t)$ について c_n を計算すると以下が得られる.

$$\begin{aligned} c_n &= \frac{1}{T_0} \int_{-T_0/2}^{T_0/2} \delta(t) e^{-jn\omega_0 t} dt \\ &= \frac{1}{T_0} e^{-jn\omega_0 \cdot 0} \\ &= \frac{1}{T_0} \end{aligned} \quad (\text{C.8})$$

このように, c_n は, n に依存せず定数となる. 議論とは関係ないため, 以下ではこの係数を 1 として扱う. 式 C.8 で得られた c_n を式 C.5 に代入すると以下が得られる.

$$\begin{aligned} x(t) &= \sum_{n=-\infty}^{\infty} e^{jn\omega_0 t} \\ &= \sum_{n=-\infty}^{\infty} e^{jnt2\pi/T_0} \end{aligned} \quad (\text{C.9})$$

次に, パルス列のフーリエ変換を示す. フーリエ変換では, 以下の式を用いて入力波形 $f(t)$ をスペクトル $F(\omega)$ に変換する.

$$F(\omega) = \int_{-\infty}^{\infty} f(t) e^{-j\omega t} dt \quad (\text{C.10})$$

周期 T_0 のパルス列 $x(t)$ を式 C.10 に代入すると以下が得られる.

$$\begin{aligned} X(\omega) &= \int_{-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \delta(t - nT_0) e^{-j\omega t} dt \\ &= \sum_{n=-\infty}^{\infty} e^{-jnT_0\omega} \\ &= \sum_{n=-\infty}^{\infty} e^{-jn2\pi\omega/\omega_0} \end{aligned} \quad (\text{C.11})$$

式 C.11 と式 C.9 より, 最終的なスペクトル $X(\omega)$ は, 以下のようになる.

$$\begin{aligned} X(\omega) &= \sum_{n=-\infty}^{\infty} e^{-jn2\pi\omega/\omega_0} \\ &= \sum_{n=-\infty}^{\infty} \delta(\omega - n\omega_0) \end{aligned} \quad (\text{C.12})$$

よって, 式 C.1 の関係は正しいことが証明された.

離散系における群遅延の導出 (3 章)

群遅延 $\tau(\omega)$ は, 位相 $\theta(\omega)$ の周波数微分に負号を加えたものとして定義されている.

$$\tau(\omega) = -\frac{d\theta(\omega)}{d\omega} \quad (\text{C.13})$$

本論文では、次のようにして群遅延の計算を行った。入力信号を $x(t)$ としたとき、入力信号のスペクトル $X(\omega)$ を以下のように表す。

$$\begin{aligned} X(\omega) &= \mathcal{F}(x(t)) \\ &= a(\omega) + jb(\omega) \end{aligned} \quad (\text{C.14})$$

\mathcal{F} はフーリエ変換を表す。スペクトル $X(\omega)$ の振幅と位相は次の式となる。

$$\begin{aligned} |X(\omega)| &= \sqrt{a^2(\omega) + b^2(\omega)} \\ \theta(\omega) &= \arctan\left(\frac{b(\omega)}{a(\omega)}\right) \end{aligned} \quad (\text{C.15})$$

このとき位相の周波数微分は以下の式のようにになる。

$$\begin{aligned} \tau(\omega) &= -\frac{d\theta(\omega)}{d\omega} \\ &= -\frac{1}{1 + \left(\frac{b(\omega)}{a(\omega)}\right)^2} \frac{d\left(\frac{b(\omega)}{a(\omega)}\right)}{d\omega} \\ &= \frac{-\frac{db(\omega)}{d\omega}a(\omega) + b(\omega)\frac{da(\omega)}{d\omega}}{a^2(\omega) + b^2(\omega)} \end{aligned} \quad (\text{C.16})$$

次に、フーリエ変換の周波数微分の法則を利用する。フーリエ変換の周波数微分の法則は以下のように定義されている。

$$\mathcal{F}(-jtx(t)) = \frac{dX(\omega)}{d\omega} = \frac{da(\omega)}{d\omega} + j\frac{db(\omega)}{d\omega} \quad (\text{C.17})$$

この法則により、以下が得られる。

$$\begin{aligned} \frac{da(\omega)}{d\omega} &= \text{real}(\mathcal{F}(-jtx(t))) \\ \frac{db(\omega)}{d\omega} &= \text{imag}(\mathcal{F}(-jtx(t))) \end{aligned} \quad (\text{C.18})$$

式 C.18 を式 C.16 に代入すると以下が得られる。

$$\tau(\omega) = \frac{-\text{real}(\mathcal{F}(-jtx(t)))a(\omega) + b(\omega)\text{imag}(\mathcal{F}(-jtx(t)))}{a^2(\omega) + b^2(\omega)} \quad (\text{C.19})$$

この式は、 $x(t)$ と、時間で重みつけた $-jtx(t)$ のスペクトルから直接群遅延が計算できることを示している。

謝辞

本学システム工学部の河原英紀教授には、指導教員として日頃から研究方針や研究内容について事細かに御指導賜りました。ここに深く感謝の意を表します。

本学システム工学部の入野俊夫教授には、提案手法に関するアイデアをいくつも頂きました。本論文で提案した技術に対する様々なご助言、博士論文執筆にも非常に丁寧に熱心なご指導を賜りました。ここに深く感謝します。

本学システム工学部の和田俊和教授には、博士論文執筆にあたり丁寧に御指導を頂きました。ここに深く感謝致します。

本学システム工学部聴覚メディア研究室研究員の高橋徹博士には、日頃から討論を通じて様々な御助言を頂きました。ここに深く感謝致します。

本学システム工学部の西村竜一助手には、多岐に渡り貴重な御助言を頂きました。ここに深く感謝致します。

名城大学理工学部の坂野秀樹講師(元本学システム工学部助手)には、第5章の論文執筆に対し大変丁寧に御指導を頂きました。ここに深く感謝致します。

また、本研究に御理解と御協力を賜った本学システム工学研究科知的計測クラスタ、および聴覚メディア研究室の諸氏に深く感謝致します。

末筆になりましたが、この研究の一部は、科学研究費補助金(特別研究員18-2150)による支援を受けて行われました。特別研究員として採用していただいた日本学術振興会に感謝致します。